Contenuto mentale stretto e largo

Introduzione — La Geometria del Significato

Vi è una domanda che la filosofia della mente e la semantica condividono senza avere ancora trovato una risposta soddisfacente, e che questa monografia si propone di affrontare con strumenti nuovi: la domanda su dove si trova il contenuto dei pensieri. Non dove si trova nel senso fisico — nel cervello, nei neuroni, nelle sinapsi — ma in quale spazio concettuale lo si debba situare, rispetto a quale struttura ontologica lo si debba descrivere, in quale relazione con il mondo lo si debba intendere. La risposta che qui si difende non appartiene ad alcuna delle famiglie di posizioni consolidate nel dibattito: non è internalista, non è esternalista nel senso standard, non è una sintesi di compromesso tra le due. È una proposta che mette in discussione la domanda stessa nella forma in cui è stata tradizionalmente posta, e propone di riformularla in un vocabolario geometrico che, a mio avviso, coglie la struttura del problema in modo più adeguato di qualsiasi framework precedente.

La tesi centrale — che chiamo *relazionalismo semantico* — può essere enunciata in forma compatta prima di essere difesa in dettaglio: il contenuto mentale non è una proprietà del soggetto cognitivo né una proprietà della relazione soggetto-mondo determinata dall'esterno, ma una struttura differenziale definita sull'intero spazio prodotto degli stati interni e degli ambienti possibili. Questa struttura ha una geometria — curvature, frontiere di sensibilità, invarianti topologici — e le proprietà semanticamente rilevanti dei pensieri sono proprietà di questa geometria, non proprietà di oggetti localizzati nell'uno o nell'altro termine della relazione mente-mondo.

Questa tesi emerge da una diagnosi critica della dicotomia stretto/largo che costituisce l'asse portante del dibattito da Putnam a Fodor, da Burge a Chalmers. Quella dicotomia, si argomenta, è fondata su un presupposto condiviso da tutte le posizioni in campo — il *dualismo di dominio*, la trattazione del soggetto e dell'ambiente come due domini ontologicamente indipendenti e pre-dati — che si rivela internamente circolare nel momento in cui si tenta di specificarlo con la precisione richiesta da una teoria. Il confine tra ciò che è interno al soggetto e ciò che è esterno non è dato antecedentemente al contenuto: è co-determinato con esso. Ogni teoria che assuma uno dei due termini come base per derivare l'altro è metodologicamente invertita.

Il relazionalismo semantico non dissolve il problema dissolvendone i termini. Al contrario: lo precisifica introducendo un formalismo che permette di riformulare le questioni tradizionali in modo più fine. La distinzione internalismo/esternalismo diventa una distinzione tra valori delle derivate parziali della funzione di contenuto — tra la sensibilità del contenuto alle variazioni interne e la sua sensibilità alle variazioni

ambientali — e perde il carattere di alternativa esclusiva: diventa una questione di misura, di distribuzione delle curvature, di struttura locale dello spazio semantico. Le leggi psicologiche vengono riqualificate come leggi su classi di equivalenza diffeomorfica della funzione di contenuto. Il riferimento singolare come proprietà della traiettoria del sistema cognitivo nello spazio prodotto. L'autoconoscenza come accesso parziale alla geometria dello spazio semantico — pieno sulla fibra verticale, limitato sulla struttura orizzontale. La comunicazione come accordo parziale tra geometrie semantiche distinte, misurabile come ampiezza e forma della regione di convergenza.

Questa ampiezza di implicazioni non è un difetto di controllo ma è strutturalmente attesa da una proposta che tocca le fondamenta. Il contenuto mentale è il concetto che connette la psicologia cognitiva alla semantica formale, l'epistemologia alla filosofia della scienza, la teoria dell'azione alla pragmatica linguistica. Una proposta che ne modifichi la struttura di base riverbera su tutti questi domini — e la monografia accompagna queste riverberazioni con la pazienza che meritano, senza abbreviarle in dichiarazioni programmatiche.

Un cenno sulla strategia argomentativa. La monografia procede in tre movimenti distinti. Il primo — le Parti Prima e Seconda, riservate alla ricostruzione storica — ha un carattere necessariamente espositivo: il relazionalismo semantico deve essere valutato rispetto al problema che si propone di risolvere, e quel problema ha una struttura che è emersa attraverso il confronto tra Putnam, Kripke, Burge, Fodor e le loro risposte reciproche. Chi conoscesse già questo dibattito in dettaglio potrebbe leggere queste parti come semplice orientamento terminologico e procedere direttamente al secondo movimento. Il secondo movimento — le Parti dalla Terza alla Decima — sviluppa la proposta positiva con il suo apparato formale, applicandola sistematicamente ai problemi che il dibattito tradizionale lasciava irrisolti. Il terzo movimento — le Parti dall'Undicesima alla Dodicesima, incluse le due obiezioni principali — sottopone la proposta a pressione: non per addolcirla ma per determinarne con precisione i limiti reali e le risorse difensive. Una teoria che non conosce i propri limiti è una teoria che non conosce se stessa.

Una precisazione sul metodo formale. Il formalismo introdotto nel corpo della monografia — la funzione di contenuto $\mathscr{C} : S \times E \to P$, le sue derivate parziali, le nozioni di curvatura e separabilità semantica, la composizione funtoriale, la regione di accordo — non pretende di essere una teoria matematica compiuta con tutti i teoremi dimostrati. Pretende di essere un vocabolario preciso che sostituisce il vocabolario impreciso con cui il dibattito stretto/largo ha lavorato fino ad ora. La precisione del vocabolario è la condizione del progresso: non perché i problemi filosofici si risolvano con i simboli, ma perché i simboli impediscono di nascondere nelle parole le ambiguità che la filosofia deve affrontare a viso aperto. L'Appendice Formale raccoglie in forma sistematica le definizioni introdotte nel corso del testo, permettendo di verificarne la coerenza interna indipendentemente dalla lettura lineare.

La tesi finale — che la mente e il mondo non siano contenitori separati tra i quali il contenuto si distribuisce, ma coordinate di uno spazio semantico curvo di cui il

contenuto è la struttura differenziale — è, nella sua forma più compatta, una proposta metafilosofica oltre che semantica. Suggerisce che la filosofia della mente abbia bisogno di un cambiamento di categoria concettuale simile a quello che la fisica ha compiuto con la geometria differenziale: dal pensare in termini di oggetti in spazi piatti al pensare in termini di strutture in spazi curvi. Se questa proposta è corretta, molte delle controversie che hanno occupato la filosofia analitica del secondo Novecento non si risolvono scegliendo un lato — non si vince adottando l'internalismo o l'esternalismo — ma si dissolvono riconoscendo che i due lati erano mal posti come alternative. Non è una posizione comoda. Ma è, si spera, una posizione più vera.

Mente, Mondo e Significato: dalla Twin Earth all'Anti-individualismo

La Rivoluzione Esternalista e l'Anatomia del Riferimento

I. Un'eredità cartesiana e il suo dissolvimento

Per almeno tre secoli la riflessione filosofica sulla mente ha operato sotto l'ipoteca di un'intuizione cartesiana tanto potente quanto, si rivelerà, insostenibile: che il contenuto degli stati mentali di un soggetto sia determinato interamente da ciò che accade *dentro* quel soggetto, dalla sua costituzione interna, dai suoi stati funzionali o neurologici, indipendentemente da qualsiasi relazione con il mondo esterno. L'idea si presenta con una certa forza pre-teoretica. Se due persone condividessero ogni singola proprietà intrinseca — ogni sinapsi identicamente configurata, ogni struttura computazionale perfettamente replicata, ogni disposizione comportamentale specularmente riprodotta — sembrerebbe naturale concludere che esse pensano le medesime cose, nutrono le medesime credenze, attribuiscono ai propri stati mentali il medesimo contenuto. Il contenuto, su questa concezione, è *stretto* (*narrow*): sopravviene sulle proprietà intrinseche del soggetto, è interamente localizzato nella scatola cranica, immune alle vicissitudini del mondo che circonda il pensatore.

Questa posizione — che prende il nome di *internalismo* semantico o psicologico, e che Jerry Fodor difenderà con estrema coerenza come condizione di possibilità di una scienza cognitiva genuina — ha radici profonde. Nel vocabolario della tradizione empirista, le idee sono oggetti mentali il cui contenuto è determinato dalla loro origine esperienziale e dalle relazioni associative che intrattengono con altre idee: un apparato che, nella sua logica profonda, rimane internal. Nel vocabolario fregeano, il *Sinn* — il senso, il modo in cui un oggetto ci è dato — è distinto dal *Bedeutung* — il riferimento, l'oggetto cui il termine punta. Il senso è, per così dire, il lato mentale o intersoggettivo del significato, mentre il riferimento appartiene al dominio estralinguistico. Un'interpretazione psicologistica di Frege, pur scorretta nei dettagli, avrebbe potuto sostenere che il contenuto del pensiero è determinato dal senso — e il senso, a

differenza del riferimento, sembra dipendere da qualcosa di interno al soggetto o alla comunità linguistica, non dall'oggetto referito in sé.

È contro questo sfondo consolidato che, nel 1975, Hilary Putnam pubblica il saggio che avrebbe ridisegnato l'intera cartografia del problema: *The Meaning of "Meaning"*. L'argomentazione putnamiana non è una semplice confutazione dell'internalismo: è qualcosa di più radicale, un tentativo di mostrare che la nozione stessa di "significato" — inteso come entità psicologicamente localizzata capace di determinare sia l'estensione di un termine sia le condizioni di correttezza di un pensiero — è fondamentalmente incoerenete. Il celebre verdetto suona come un aforisma: *"Cut the pie any way you like, 'meanings' just ain't in the head."* Ma prima di valutare questo verdetto occorre comprendere la macchina argomentativa che lo produce.

II. La Terra Gemella e la struttura dell'esperimento mentale

L'esperimento mentale della Terra Gemella (*Twin Earth*) è diventato talmente pervasivo nel dibattito filosofico da rischiare di essere conosciuto per fama senza essere compreso nella sua architettura logica. Vale dunque la pena ricostruirlo con la lentezza che merita.

Putnam invita a immaginare un pianeta — la Terra Gemella, appunto — che sia in tutto e per tutto identico alla Terra eccetto per un fatto: il liquido che su quel pianeta scorre nei fiumi, riempie gli oceani, viene bevuto dagli abitanti e chiamato con la parola "water" (nella versione anglofona dell'esperimento) *non è* H_2O. Il liquido ha la medesima apparenza macroscopica dell'acqua terrestre — è incolore, insapore, inodore, disseta, spegne incendi, evapora a 100°C a pressione atmosferica standard — ma la sua struttura molecolare è radicalmente diversa, designabile con la sigla convenzionale XYZ. Questa differenza è, nel 1750, completamente inaccessibile agli abitanti di entrambi i pianeti, la cui chimica è ancora prescientifica e incapace di distinguere i due liquidi a livello subvisibile.

Consideriamo ora Oscar e il suo doppio perfetto sulla Terra Gemella, chiamato Oscar*. I due soggetti sono intrinsecamente indistinguibili: ogni atomo del loro corpo occupa la medesima posizione, ogni stato neuronale è identicamente configurato, ogni disposizione comportamentale è una copia speculare dell'altra. Quando Oscar usa il termine "water" e Oscar* usa la parola corrispondente nella propria lingua, i due sembrano essere nel medesimo stato psicologico interno. Eppure — e qui risiede il cuore dell'argomento — il termine "water" nella bocca di Oscar *si riferisce* ad H_2O, mentre il termine corrispondente nella bocca di Oscar* si riferisce ad XYZ. Le *estensioni* dei due termini sono distinte: insiemi di oggetti radicalmente diversi cadono sotto ciascuno di essi. Poiché le proprietà intrinseche dei due parlanti sono identiche, e poiché le estensioni sono diverse, ne segue che l'estensione — e dunque almeno una componente cruciale del significato — non è determinata dalle proprietà intrinseche

del parlante. Il significato, in questa accezione, è *largo* (*wide*): dipende dalla relazione del soggetto con il suo ambiente.

La struttura logica dell'argomento può essere resa esplicita nella forma seguente. Sia S(x) il predicato "x è nello stesso stato psicologico interno di y" per soggetti x, y. Sia Ext(t, x) la funzione che assegna al termine t come usato dal parlante x la sua estensione — l'insieme degli oggetti cui t si applica correttamente. L'internalismo semantico sostiene la seguente tesi di sopravvenienza:

$$\forall x \, \forall y \, [S(x, y) \rightarrow \text{Ext}(t_x, x) = \text{Ext}(t_y, y)]$$

dove t_x e t_y sono i termini corrispondenti nei due idioletti. La Terra Gemella è costruita precisamente per falsificare questo schema: Oscar e Oscar* soddisfano l'antecedente (sono in tutto e per tutto psicologicamente identici) ma non il conseguente (le estensioni di "water" e del suo corrispondente gemello sono disgiunte). La tesi di sopravvenienza internalista è dunque falsa, stando all'esperimento mentale.

III. Termini per tipi naturali e la divisione del lavoro linguistico

L'argomento putnamiano non si limita all'esperimento mentale della Terra Gemella: esso si inserisce in una teoria più ampia del significato dei *termini per tipi naturali* (*natural kind terms*). Questi termini — "acqua", "oro", "tigre", "limone" — hanno, secondo Putnam, una struttura semantica peculiare che li distingue radicalmente dai termini definibili per via di condizioni necessarie e sufficienti espresse in termini osservativi.

La visione tradizionale, che Putnam ricollega alla teoria descrittivista del significato associata a Russell e al primo Wittgenstein nella sua interpretazione corrente, sosteneva che il significato di un termine è un fascio di descrizioni che il parlante associa ad esso. Il significato di "acqua" sarebbe qualcosa come: *il liquido incolore, insapore, che disseta, che scorre nei fiumi, ecc.* L'estensione del termine sarebbe allora determinata da questo fascio descrittivo: appartengono all'estensione di "acqua" tutti e soli gli oggetti che soddisfano le descrizioni associate. Su questa concezione, il significato è pienamente interno al soggetto — è la rete di credenze e associazioni che questi intrattiene con il termine.

Putnam demolisce questa concezione su più fronti. Il primo attacco riguarda la *flessibilità referenziale*: i parlanti ordinari usano correttamente termini per tipi naturali senza essere in grado di fornire descrizioni distintive adeguate. La maggior parte dei parlanti non sa distinguere un olmo da un faggio; eppure "olmo" ed "elmtree" si riferiscono a specie diverse, e un parlante che dica "l'olmo davanti a casa mia" si riferisce all'albero giusto anche se non potrebbe discriminarlo da un faggio. Ciò che consente la riferimento corretto non è il fascio descrittivo posseduto individualmente,

ma la catena causale che lega l'uso del termine all'esemplare originario della specie, e la *divisione del lavoro linguistico* che delega agli esperti — botanici, chimici, zoologi — la determinazione delle condizioni di correttezza del termine.

Formalmente, Putnam introduce la nozione di *indice* (*index*) del significato: il significato di un termine non è una singola entità psicologica ma una quadrupla:

⟨ marcatori sociolinguistici, marcatori semantici, condizione stereotipica, estensione ⟩

dove i marcatori sociolinguistici indicano se il termine appartiene al vocabolario degli esperti o del profano, i marcatori semantici sono predicati del genere "è un solido", "è un liquido", "è un animale", la condizione stereotipica cattura le proprietà tipiche associate all'uso ordinario, e l'estensione è l'insieme effettivo degli oggetti cui il termine si applica. Crucialmente, l'estensione non è *derivata* dalle altre componenti: essa è determinata da come il mondo è fatto, non da come i parlanti lo concepiscono. L'acqua è H_2O indipendentemente da ciò che chiunque — prima della chimica moderna — sapeva o credeva dell'acqua.

Questa tesi si connette direttamente alla teoria causale del riferimento elaborata in parallelo da Saul Kripke in *Naming and Necessity* (le conferenze del 1970, pubblicate nel 1972 e poi nel 1980). Per Kripke, i nomi propri sono *designatori rigidi*: designano lo stesso individuo in tutti i mondi possibili in cui quell'individuo esiste, a prescindere da quali proprietà quell'individuo possieda in ciascun mondo. La connessione tra il nome e l'oggetto nominato non è mediata da un fascio di descrizioni, ma da una *catena causale storica* che risale all'atto battesimale originario — il momento in cui il nome viene conferito all'oggetto, tipicamente in presenza dell'oggetto stesso o di descrizioni individuanti. Putnam estende questa idea ai termini per tipi naturali: "acqua" designa rigidamente H_2O perché è stato introdotto in riferimento a esemplari di H_2O — l'atto battesimale originario —, e questa connessione causale si trasmette lungo la catena dei successivi usi del termine.

La convergenza tra Kripke e Putnam definisce quello che verrà chiamato l'*esternalismo semantico* — o, nella terminologia di Burge, *anti-individualismo*. Il significato, il contenuto, il riferimento di un termine non sono determinati da ciò che è interno alla mente del parlante: dipendono dalla natura del mondo e dalla storia causale che lega il parlante a quel mondo.

IV. Il contenuto largo e le sue implicazioni per la teoria della mente

L'esternalismo semantico, nei termini in cui è stato fin qui presentato, potrebbe sembrare confinato alla filosofia del linguaggio: una teoria sul riferimento dei termini, non necessariamente una teoria sugli stati mentali. Ma Putnam è esplicito nel sostenere che l'argomento si applica direttamente ai contenuti dei pensieri. Se il significato di

"acqua" non è nella testa, allora neppure il *contenuto* del pensiero *che l'acqua disseta* è interamente nella testa. Oscar e Oscar* non pensano la medesima cosa quando pensano "l'acqua è bagnata": Oscar pensa qualcosa riguardo ad H_2O, Oscar* pensa qualcosa riguardo ad XYZ. I due pensieri hanno *contenuti* diversi — sono, in un senso preciso, pensieri *su* oggetti diversi — benché siano prodotti da stati cerebrali intrinsecamente identici.

Questa conclusione ha conseguenze di vasta portata per la teoria della mente e per la psicologia cognitiva. La *psicologia del senso comune* — la cosiddetta *folk psychology* — attribuisce agli agenti stati intenzionali (credenze, desideri, intenzioni, speranze) caratterizzati da contenuti proposizionali. Quando spiego il comportamento di Mario dicendo che egli credeva che il treno partisse alle otto, attribuisco a Mario una credenza con un certo contenuto. Ma se il contenuto degli stati mentali è largo — se dipende non solo dagli stati interni del soggetto ma dalle sue relazioni causali con il mondo —, allora due soggetti intrinsecamente identici possono avere credenze con contenuti diversi. Le leggi psicologiche formulate in termini di contenuti larghi non sarebbero, almeno in linea di principio, leggi sovrapponibili alle leggi che descrivono gli stati interni.

Fodor, come si vedrà nella seconda parte di questo studio, risponderà a questa sfida sostenendo che una psicologia cognitiva scientificamente rispettabile deve fare a meno del contenuto largo e lavorare esclusivamente con il contenuto stretto — quello che sopravviene sugli stati interni e che dunque può figurare in generalizzazioni psicologiche universalmente quantificate su soggetti intrinsecamente identici. Ma questa risposta, pur attraente, incontra difficoltà serie: come si definisce il contenuto stretto senza ricorrere al contenuto largo? E ha il contenuto stretto una realtà psicologica genuina o è un costrutto artificiale senza valore esplicativo autonomo?

Queste domande strutturano il dibattito che si estenderà per i decenni successivi e che questa monografia si propone di seguire fino alle sue diramazioni più tecniche.

L'Anti-individualismo di Burge, la Risposta Internalista e l'Architettura Formale della Sopravvenienza

---

V. Tyler Burge e la dimensione sociale del contenuto

Se l'esternalismo putnamiano trae la propria forza dalla struttura fisico-chimica del mondo — dalla natura di H_2O, dall'essenza dei tipi naturali —, la variante proposta da Tyler Burge percorre un itinerario diverso e, per certi versi, ancora più radicale. Burge mostra che il contenuto degli stati mentali dipende non soltanto dall'ambiente fisico

del soggetto, ma dal suo ambiente *sociale* e istituzionale: dalle pratiche linguistiche della comunità entro cui egli opera, dalle norme semantiche che quella comunità ha sedimentato, dalle definizioni che gli esperti custodiscono e che il parlante ordinario non possiede né è tenuto a possedere. Questo *esternalismo sociale* — che Burge preferisce chiamare *anti-individualismo* per sottolineare il contrasto con la tradizione individualista da Cartesio a Fodor — è argomentato mediante una serie di esperimenti mentali pubblicati nel saggio fondamentale del 1979, *Individualism and the Mind*.

Il più celebre di questi esperimenti riguarda l'artrite. Supponiamo che Tizio abbia sviluppato nel tempo una credenza falsa: crede di avere l'artrite alla coscia, mentre l'artrite — per definizione medica — è un'infiammazione delle articolazioni e non può manifestarsi nella coscia. Il termine "artrite" nel dialetto mentale di Tizio ha dunque un'applicazione aberrante: egli lo usa per coprire sia i casi corretti (dolori articolari) sia un caso incorretto (dolore muscolare alla coscia). Ciononostante, molte delle sue credenze sull'artrite sono vere e correttamente espresse: sa che l'artrite è dolorosa, che si tratta di una condizione infiammatoria, che si cura con certi farmaci. Burge sostiene che, in questo scenario, Tizio ha autenticamente una credenza espressa con il pensiero *"ho l'artrite alla coscia"* — una credenza falsa, ma genuina, il cui contenuto è determinato dal significato sociale del termine "artrite" nella sua comunità linguistica.

Viene poi introdotto il soggetto controfattuale: una persona intrinsecamente identica a Tizio — medesimi stati neurali, medesima storia causale con i dolori fisici, medesime disposizioni comportamentali — ma inserita in una comunità linguistica alternativa, dove il termine "artrite" (o il suo equivalente) si applica non solo alle infiammazioni articolari ma anche a certi dolori muscolari. In questa comunità controfattuale, l'uso del termine fatto dal soggetto non sarebbe aberrante: il suo stato di credenza sarebbe corretto rispetto alle norme semantiche locali. Burge conclude che i contenuti delle credenze del soggetto nel caso reale e in quello controfattuale sono *diversi*: nel primo caso, la credenza ha il contenuto *"ho l'artrite alla coscia"* dove "artrite" significa ciò che la medicina stabilisce; nel secondo, un contenuto diverso dove il termine ha un'estensione più ampia.

La differenza di contenuto si produce nonostante l'identità intrinseca dei due soggetti. La struttura dell'argomento riproduce quella putnamiana: si viola la tesi di sopravvenienza internalista, che ora può essere espressa così:

$$\forall x \, \forall y \, [\, I(x, y) \rightarrow C(b_x) = C(b_y) \,]$$

dove $I(x, y)$ significa che x e y sono intrinsecamente identici (stessa storia neurofisiologica, stessa serie di esperienze sensoriali), b_x e b_y sono le credenze corrispondenti nei due soggetti, e C è la funzione che assegna a ciascuna credenza il suo contenuto proposizionale. L'esperimento dell'artrite esibisce una coppia $\langle x, y \rangle$ tale che $I(x, y)$ ma $C(b_x) \neq C(b_y)$: il soggetto reale ha la credenza *che ha l'artrite alla coscia*, con contenuto falso; il soggetto controfattuale ha una credenza di contenuto diverso, non classificabile come credenza sull'artrite nel senso medico del termine.

La dimensione aggiuntiva rispetto a Putnam è significativa. Nell'esperimento della Terra Gemella, la differenza di contenuto era riconducibile alla struttura chimica del mondo: H_2O versus XYZ erano entità fisicamente distinte. Nell'esperimento dell'artrite, la differenza di contenuto dipende interamente da convenzioni sociali e istituzionali: non c'è alcun fatto fisico profondo che distingua il dolore muscolare dall'infiammazione articolare ai fini della classificazione semantica; è la pratica medica e la norma linguistica che tracciano il confine. Il contenuto mentale è dunque sensibile non soltanto all'ambiente fisico ma all'ambiente normativo — alla rete di pratiche, accordi e autorità epistemiche entro cui il soggetto è immerso.

Burge estende l'analisi a numerosi altri casi: i termini tecnici della giurisprudenza, le categorie della fisica, i nomi delle malattie. In ognuno di questi domini, un parlante che usi correttamente — nel senso pragmatico dell'integrazione comunicativa — un termine specialistico senza padroneggiarne la definizione tecnica ha comunque credenze i cui contenuti sono determinati da quella definizione tecnica e non da ciò che egli individualmente sa. Il contenuto mentale, su questa concezione, è *irriducibilmente sociale*: appartiene a una dimensione terza tra l'individuo e il mondo fisico, la dimensione delle pratiche normative condivise.

VI. L'internalismo foderiano e il contenuto stretto: un programma di salvezza

Jerry Fodor accoglie le conclusioni esternaliste con una consapevolezza acuta ma risponde proponendo una distinzione che dovrebbe, nelle sue intenzioni, preservare l'autonomia della psicologia cognitiva: la distinzione tra *contenuto largo* (*wide content*) e *contenuto stretto* (*narrow content*). Il contenuto largo è ciò che gli argomenti di Putnam e Burge descrivono: il contenuto degli stati mentali determinato dalle relazioni causali e normative del soggetto con l'ambiente fisico e sociale. Il contenuto stretto è ciò che sopravviene esclusivamente sugli stati interni: è quel residuo di "contenuto" che rimane invariante quando si varia sistematicamente l'ambiente lasciando ferma la costituzione interna.

Fodor è esplicito sul perché la psicologia cognitiva abbia bisogno del contenuto stretto. La psicologia — come qualsiasi scienza speciale — ambisce a formulare generalizzazioni nomologiche, leggi ceteris paribus che connettono tipi di stati mentali a tipi di comportamenti. Una generalizzazione del tipo "chiunque creda che P e voglia che Q e ritenga che fare A produca Q, tenderà a fare A" deve essere universalmente applicabile: deve valere per Oscar e per Oscar*, per qualsiasi coppia di soggetti intrinsecamente identici indipendentemente dall'ambiente in cui si trovano. Se le leggi psicologiche fossero formulate in termini di contenuto largo, varrebbero per Oscar ma non necessariamente per Oscar*, poiché i contenuti larghi dei due differiscono. Ciò renderebbe impossibile l'unificazione nomologica che caratterizza una scienza genuina.

La soluzione foderiana, elaborata soprattutto in *Psychosemantics* (1987) e poi in *A Theory of Content* (1990), è di separare nettamente la dimensione psicologica — governata dal contenuto stretto — dalla dimensione semantica — governata dal contenuto largo. La psicologia cognitiva studia la sintassi delle rappresentazioni mentali e le leggi che la governano; la semantica studia come tali rappresentazioni si collegano al mondo esterno. Le due discipline sono complementari ma metodologicamente indipendenti.

Il problema è che il contenuto stretto, così concepito, è terribilmente difficile da caratterizzare positivamente. Non è semplicemente la forma sintattica della rappresentazione — Fodor non è un sintatticista puro, e riconosce che le rappresentazioni hanno un contenuto genuino, non soltanto una struttura formale. Ma che cos'è, allora, il contenuto stretto? Fodor propone due caratterizzazioni diverse in momenti diversi della sua riflessione.

La prima caratterizzazione fa appello ai *ruoli computazionali* delle rappresentazioni: il contenuto stretto di una rappresentazione è determinato dal suo ruolo nella rete di inferenze e transizioni computazionali che il sistema cognitivo compie. Questa concezione ha parentele con la *semantica del ruolo concettuale* (*conceptual role semantics*) di Block e Harman: il contenuto è determinato dalla posizione della rappresentazione nello spazio inferenziale. Ma Fodor è critico verso questa identificazione perché teme che il ruolo computazionale sia troppo olistico — dipenda dall'intera rete di credenze del soggetto — e che ciò metta a rischio la possibilità di leggi psicologiche stabili.

La seconda caratterizzazione, più ambiziosa e tecnicamente articolata, fa appello ai *mondi nozionali* (*notional worlds*). L'idea è che ogni soggetto, in virtù della propria costituzione interna, sia "tarato" su un certo insieme di mondi possibili: i mondi in cui i suoi stati interni avrebbero contenuto largo corretto, i mondi in cui l'ambiente esterno corrisponderebbe alla rappresentazione interna. Il contenuto stretto sarebbe allora una *funzione* dagli ambienti ai contenuti larghi: non un contenuto fisso, ma un contenuto *relativo all'ambiente*. Formalmente, se W denota l'insieme dei mondi possibili e C_L(s, w) è il contenuto largo dello stato s nel mondo w, allora il contenuto stretto di s può essere caratterizzato come la funzione:

$$cs(s) : W \rightarrow \text{Prop}$$
$$cs(s)(w) = C_L(s, w)$$

dove Prop è l'insieme delle proposizioni. Il contenuto stretto è dunque un *contenuto-funzione*, una mappa che associa a ciascun contesto ambientale il contenuto largo che lo stato avrebbe in quel contesto. Oscar e Oscar* hanno lo stesso contenuto stretto per "water" — la medesima funzione dagli ambienti alle proposizioni — ma contenuti larghi diversi, perché i loro ambienti effettivi differiscono.

Questa caratterizzazione è elegante, ma si espone a obiezioni di varia natura. La più immediata è che il contenuto stretto, così inteso, non è *un* contenuto ma una molteplicità di contenuti possibili organizzata in una funzione: un oggetto matematico astratto piuttosto che un contenuto psicologicamente reale. Come può un tale oggetto svolgere il ruolo esplicativo che la psicologia le assegna? Quando spiego che Mario ha

aperto il frigorifero perché credeva che ci fosse della birra, non sto attribuendo a Mario una funzione dagli ambienti alle proposizioni: sto attribuendo una credenza con un contenuto determinato. Il contenuto stretto, nella versione foderiana, sembra troppo astratto per fungere da elemento genuino della spiegazione psicologica intenzionale.

VII. La sopravvenienza locale e il problema del livello di granularità

Una delle difficoltà più sottili che affliggono il dibattito stretto/largo riguarda la *granularità* della sopravvenienza. Quando si dice che il contenuto stretto sopravviene sugli stati interni del soggetto, a quale livello di descrizione ci si riferisce? Sopravviene sui neuroni? Sugli stati funzionali di livello intermedio? Sui pattern di attivazione distribuiti? La risposta a questa domanda non è neutra, e influisce in modo decisivo sulla plausibilità dell'internalismo.

Consideriamo la seguente formalizzazione. Sia $\Sigma = \{\sigma_1, \sigma_2, ..., \sigma\}$ l'insieme delle proprietà intrinseche fisicamente rilevanti di un soggetto, descritte a un certo livello di granularità g. La tesi di sopravvenienza internalista al livello g afferma:

$$\forall x \, \forall y \, [\, (\forall \sigma_i \in \Sigma : \sigma_i(x) \leftrightarrow \sigma_i(y)) \rightarrow C(x) = C(y) \,]$$

dove C(x) indica il contenuto degli stati mentali di x. Ora, il problema è che il livello di granularità fisicamente appropriato per la psicologia cognitiva non è ovviamente determinato. Se si scende al livello subatomico, la distinzione tra stati interni e relazioni con l'esterno rischia di collassare: le proprietà fisiche delle particelle costituenti il cervello di Oscar dipendono, in ultima istanza, dall'intero campo gravitazionale ed elettromagnetico dell'universo, incluso il campo generato da H_2O nei fiumi terrestri. Una sopravvenienza puramente locale non è garantita neppure a livello fisico elementare.

Più concretamente: esiste un *raggio causale* entro cui gli stati fisici di un sistema dipendono dall'ambiente circostante. Se questo raggio include — anche indirettamente, attraverso catene causali lunghissime — la struttura molecolare dell'acqua terrestre, allora la distinzione tra "stati interni" e "relazioni con l'esterno" è una distinzione di convenienza teorica, non un confine ontologico netto. L'internalista potrebbe rispondere che ciò che conta è la sopravvenienza *pelle*, cioè sui soli stati fisici della superficie del corpo — la pelle, appunto, come confine convenzionale tra il soggetto e il mondo. Ma questa risposta è arbitraria: perché la pelle e non il cervello? Perché il cervello e non il sistema nervoso periferico? Ogni scelta del confine è teoricamente motivata, non naturalmente data.

La difficoltà si acuisce quando si consideri il fenomeno della *estesa cognizione* (*extended cognition*), proposta da Andy Clark e David Chalmers nel fondamentale articolo *The*

Extended Mind del 1998. Clark e Chalmers argomentano che in certi casi il confine cognitivamente rilevante non coincide né con la pelle né con il cranio: quando un soggetto usa un taccuino per compensare un deficit di memoria, o usa uno smartphone per reperire informazioni, questi strumenti esterni svolgono una funzione cognitiva che in altri soggetti è svolta da processi neurali interni. Il *principio di parità* — se una funzione cognitiva svolta da un artefatto esterno avrebbe, se svolta internamente, contato come un processo cognitivo, allora conta come processo cognitivo anche quando è esterna — suggerisce che il confine del sistema cognitivo sia funzionalmente determinato, non anatomicamente fisso. Se l'esternalismo semantico di Putnam e Burge si estende in questa direzione, il contenuto mentale diventa non soltanto sensibile all'ambiente ma potenzialmente *distribuito* su di esso: una conseguenza che l'internalismo foderiano non può assorbire senza trasformarsi in qualcosa di radicalmente diverso.

VIII. Il dibattito sulla compatibilità: contenuto largo e spiegazione causale

Una delle tensioni più produttive del dibattito riguarda la relazione tra contenuto largo e *spiegazione causale del comportamento*. L'argomento più pressante contro l'esternalismo è stato formulato con precisione da Fred Dretske e, in forma diversa, da Fodor stesso: se il contenuto largo di uno stato mentale dipende dall'ambiente esterno, come può quel contenuto figurare nella spiegazione causale del comportamento? La causa del comportamento deve essere qualcosa di *localmente* presente al momento dell'azione; ma il contenuto largo di una credenza dipende da eventi passati (la storia causale che lega il soggetto al referente) e da entità esterne (la struttura molecolare dell'acqua, le norme semantiche della comunità). Come può qualcosa di storicamente e geograficamente distante causare, qui e ora, il movimento di un corpo?

Questo argomento ha la struttura di una variante del *problema del contenuto causalmente inerte*: se il contenuto largo non è localmente presente, esso non può causare il comportamento, e allora la psicologia intenzionale — che spiega il comportamento in termini di contenuti — è o falsa o irriducibile a una psicologia causale genuina. Si delinea il fantasma dell'epifenomenismo: il contenuto sarebbe un epifenomeno causalmente inerte sovrapposto agli stati fisici, i quali soli eserciterebbero il vero potere causale.

Jaegwon Kim ha formalizzato questo timore nella sua critica al dualismo della proprietà: se le proprietà mentali sopravvengono su quelle fisiche senza essere identiche ad esse, e se le proprietà fisiche chiudono causalmente il dominio fisico, allora le proprietà mentali non possono esercitare un potere causale autonomo senza produrre una sovraddeterminazione sistematica. Lo schema è il seguente. Sia M una proprietà mentale (diciamo: credere che vi sia acqua nella brocca), P la proprietà fisica su cui M sopravviene, e B l'evento comportamentale prodotto (prendere la brocca). Se P è causalmente sufficiente per B, e M sopravviene su P, allora M non aggiunge nulla

causalmente: M è per B ciò che il colore di un calcio è per l'effetto del calcio — una proprietà accompagnatoria causalmente inerte.

La risposta esternalista più elaborata proviene da filosofi come Louise Antony e da Burge stesso: essi sostengono che il problema dell'esclusione causale di Kim si basa su un'ontologia causale troppo restrittiva, che privilegia le cause di tipo "evento fisico puntuale" e non lascia spazio alle cause di tipo "proprietà di livello superiore". Se si adotta un'ontologia causale pluralista — se si ammette che spiegazioni a livelli diversi di descrizione possano essere egualmente valide e non competitive —, allora il contenuto largo può figurare in spiegazioni causali genuine senza competere con le spiegazioni fisiche sottostanti. La questione, nelle sue implicazioni più profonde, tocca problemi generali della filosofia della scienza: la riduzione, l'emergenza, la struttura delle scienze speciali. Non è risolvibile, quindi, soltanto con risorse semantiche, ma richiede una presa di posizione sull'architettura generale del reale.

Un'altra linea di risposta, seguita tra gli altri da Stephen Yablo con la sua teoria della *causazione-per-esclusione* e da Karen Neander nell'ambito delle teorie teleosemantiche, è di ridefinire la nozione di contenuto in modo da integrare la dimensione causale fin dalla base: il contenuto di uno stato mentale sarebbe determinato dalla sua funzione biologica, dall'ambiente evolutivo di adattamento in cui la capacità rappresentazionale si è selezionata. Questa prospettiva — che verrà esaminata nella parte terza di questo studio — promette di riconciliare il carattere esterno del contenuto con la sua efficacia causale locale, ma apre a sua volta nuovi problemi di non minore complessità.

La Dicotomia come Artefatto: Verso un Relazionalismo Semantico

---

IX. La circolarità fondativa della distinzione

Il dibattito tra internalismo ed esternalismo semantico ha prodotto, nel corso di cinquant'anni, una letteratura di straordinaria sottigliezza tecnica. Eppure vi è, a mio avviso, una difficoltà strutturale che nessuna delle posizioni in campo ha affrontato con sufficiente radicalità: la distinzione stessa tra contenuto stretto e contenuto largo è fondata su una nozione di "confine del soggetto" che si rivela, a un'analisi ravvicinata, internamente circolare. Questa circolarità non è un difetto riparabile per aggiustamenti locali alla teoria: è un sintomo di un'impostazione ontologica sbagliata, che tratta il soggetto e l'ambiente come due domini ontologicamente indipendenti e pre-dati, tra i quali il contenuto mentale dovrebbe poi distribuirsi. Propongo di chiamare questa impostazione *dualismo di dominio*, e di mostrare che la coppia concettuale stretto/largo ne è una conseguenza, non un'analisi neutrale dei fatti semantici.

Il dualismo di dominio opera, nella letteratura, come un'assunzione di sfondo raramente tematizzata. Quando Fodor afferma che il contenuto stretto sopravviene

sugli stati *interni* del soggetto, presuppone che vi sia un confine netto e ontologicamente significativo tra ciò che è interno e ciò che è esterno. Quando Putnam afferma che il significato non è nella testa, presuppone anch'egli questo confine, invertendone semplicemente il peso esplicativo: ciò che conta, per il significato, non è dentro la testa ma fuori. Entrambe le posizioni condividono il presupposto strutturale che il soggetto cognitivo sia un sistema individuato in modo indipendente dal contenuto dei propri stati mentali, e che l'ambiente sia ciò che rimane una volta che il soggetto è stato ritagliato.

Questo presupposto è falso, e la falsità si manifesta non appena si tenti di specificarlo con la precisione richiesta da una teoria. Qual è il confine del soggetto cognitivo? Le risposte disponibili sono tre: il confine anatomico (la pelle, o il cranio), il confine funzionale (il sistema che svolge le funzioni cognitive rilevanti), il confine evolutivo (il sistema selezionato per svolgere certe funzioni in un certo ambiente di adattamento). Ciascuna di queste specificazioni presuppone, nella sua formulazione, decisioni su *quali* processi contano come cognitivi e *quali* relazioni con l'ambiente sono rilevanti — decisioni che non possono essere prese senza già avere una nozione di contenuto mentale. Il confine del soggetto non è dato indipendentemente dal contenuto: è co-determinato con esso.

Si consideri il caso più semplice: il confine anatomico. Perché il cranio dovrebbe essere il confine rilevante per la determinazione del contenuto? La risposta implicita è che il cervello è il substrato dei processi cognitivi. Ma se il cervello è rilevante in quanto substrato *cognitivo*, allora la sua rilevanza dipende già da una teoria di quali processi siano cognitivi — e questa teoria, per non essere circolare, non può a sua volta fare appello al cranio come confine. Peggio ancora: vi sono casi, come quello della percezione visiva estesa, in cui la catena causale che genera la rappresentazione include non solo i processi retinici e corticali ma le proprietà fisiche dell'oggetto distale, le condizioni di illuminazione, persino le convenzioni culturali che determinano l'interpretazione dello stimolo ambiguo. Decidere dove "taglia" il processo cognitivo è già una decisione teorica carica di contenuto.

Il confine funzionale eredita il problema in forma acuta. Clark e Chalmers hanno mostrato che l'individuo cognitivo può estendersi, funzionalmente parlando, ben oltre la pelle. Ma il *principio di parità* che giustifica questa estensione — "se una funzione svolta esternamente, svolta internamente, conterebbe come cognitiva, allora conta come cognitiva anche esternamente" — è esso stesso applicabile solo a partire da una determinazione di quali funzioni siano cognitive, il che presuppone una nozione di contenuto. Si cade in una regressione: per tracciare il confine funzionale del soggetto cognitivo, ho bisogno di una teoria del contenuto; per sviluppare una teoria del contenuto, ho bisogno di sapere quali processi e quali relazioni con l'ambiente sono rilevanti; per saperlo, ho bisogno di tracciare il confine del soggetto cognitivo.

La circolarità non è logicamente viziosa nel senso di una petizione di principio: non si tratta di un argomento che assume ciò che deve dimostrare. È qualcosa di più profondo: è il segnale che soggetto e contenuto sono concetti *mutuamente costitutivi*, non ontologicamente indipendenti. Ogni teoria che assuma uno come base per derivare

l'altro è metodologicamente invertita. La distinzione stretto/largo — che chiede quanta parte del contenuto dipende da ciò che è dentro e quanta da ciò che è fuori — è mal posta perché il "dentro" e il "fuori" non sono dati antecedentemente al contenuto.

X. Verso un relazionalismo semantico: il contenuto come struttura differenziale

La diagnosi della circolarità non è sterile: indica la direzione di una proposta alternativa. Se il soggetto e l'ambiente sono mutuamente costitutivi rispetto al contenuto, allora il contenuto non è una proprietà del soggetto (internalismo) né una proprietà della relazione soggetto-mondo determinata dall'esterno (esternalismo nel senso di Putnam e Burge). Il contenuto è una struttura *irreducibilmente relazionale*: un'entità che vive nello spazio della relazione tra un sistema cognitivo e un ambiente, senza essere riducibile a nessuno dei due termini considerati isolatamente. Chiamo questa posizione *relazionalismo semantico*.

Per articolare il relazionalismo semantico con la precisione che merita occorre un apparato formale che vada oltre quello disponibile nel dibattito tradizionale. Propongo di concettualizzare il contenuto come una struttura differenziale su uno spazio prodotto, seguendo un'analogia con la geometria differenziale che non è meramente illustrativa ma coglie qualcosa di genuino sulla natura del contenuto.

Sia S lo spazio degli stati interni possibili di un sistema cognitivo — un insieme che comprende, almeno idealmente, tutti i pattern di attivazione, disposizioni funzionali, o qualsiasi altro tipo di stati interni che la teoria voglia individuare come rilevanti. Non è necessario specificare la natura fisica di questi stati: si tratta di uno spazio astratto. Analogamente, sia E lo spazio degli ambienti possibili — un insieme di situazioni fisiche, sociali e normative entro cui il sistema può essere inserito. Sia infine P l'insieme delle proposizioni (o contenuti proposizionali), inteso nel senso più neutro possibile: funzioni caratteristiche da insiemi di mondi possibili, o qualsiasi altro trattamento formale si preferisca.

Una *funzione di contenuto* è una mappa:

$$\mathscr{C} : S \times E \to P$$
$$\mathscr{C}(s, e) = p$$

che assegna a ogni coppia ⟨stato-interno, ambiente⟩ un contenuto proposizionale. Questa funzione non è la semplice unione dell'internalismo e dell'esternalismo: non dice che il contenuto è la somma di un contributo interno e di uno esterno. Dice che il contenuto è *definito sull'intero prodotto* $S \times E$, senza che uno dei due fattori abbia priorità ontologica sull'altro.

La distinzione stretto/largo trova ora una formulazione precisa come condizione sulle derivate parziali di $\mathscr{C}$. Il contenuto stretto corrisponderebbe all'ipotesi che $\mathscr{C}$ sia costante rispetto alla variazione di e mantenendo s fisso, cioè:

$\partial\mathscr{C}/\partial e = 0$ (uniformemente su $S \times E$)

Il contenuto largo puro, per contro, corrisponderebbe all'ipotesi che $\mathscr{C}$ sia costante rispetto alla variazione di s mantenendo e fisso:

$\partial\mathscr{C}/\partial s = 0$

Entrambe le condizioni, nella loro formulazione globale, sono epistemicamente insostenibili: non vi è ragione di credere che $\mathscr{C}$ sia localmente costante rispetto a nessuno dei due fattori su tutto lo spazio $S \times E$. Ciò che il relazionalismo semantico afferma è che $\mathscr{C}$ è genuinamente *bidirezionalmente sensibile*: al variare di s con e costante, il contenuto varia; al variare di e con s costante, il contenuto varia. Il contenuto non è una proprietà del punto $(s, e) \in S \times E$, ma una proprietà della *struttura locale* di $\mathscr{C}$ intorno a quel punto — e questa struttura locale è catturata dal *differenziale totale*:

$d\mathscr{C} = (\partial\mathscr{C}/\partial s)\, ds + (\partial\mathscr{C}/\partial e)\, de$

dove ds e de rappresentano variazioni infinitesime nello stato interno e nell'ambiente rispettivamente. Ogni variazione del contenuto è, in linea di principio, attribuibile a una combinazione di variazioni interne e variazioni ambientali, nella proporzione determinata dai coefficienti del differenziale in quel punto.

Questa struttura ha una conseguenza filosoficamente importante: l'internalismo e l'esternalismo non sono posizioni *alternative* ma *casi limite* della struttura relazionale. L'internalismo è il caso in cui $\partial\mathscr{C}/\partial e \approx 0$ localmente — in cui, nella regione dello spazio $S \times E$ pertinente, le variazioni ambientali producono variazioni di contenuto trascurabili rispetto alle variazioni interne. L'esternalismo è il caso opposto, $\partial\mathscr{C}/\partial s \approx 0$ localmente. La realtà semantica è, nel caso generale, una struttura dove entrambe le derivate parziali sono non nulle, e la questione empiricamente e filosoficamente interessante è la forma *locale* del differenziale: quali variazioni di quale tipo producono variazioni di contenuto, e in quale misura?

XI. L'individuazione del soggetto come problema interno alla semantica

Il relazionalismo semantico permette di riformulare il problema dell'individuazione del soggetto in termini strutturalmente più onesti. Non si tratta più di trovare un confine

pre-dato tra interno ed esterno: si tratta di determinare, per ciascuna funzione di contenuto $\mathscr{C}$, quale struttura dello spazio $S \times E$ è semanticamente rilevante.

Introduco a questo scopo la nozione di *separabilità semantica*. Diciamo che $\mathscr{C}$ è *s-separabile* rispetto a una partizione dell'ambiente $E = E_1 \cup E_2$ se:

$$\forall s \in S \; \forall e_1 \in E_1 \; \forall e_2 \in E_2 : \mathscr{C}(s, e_1) = \mathscr{C}(s, e_2)$$

cioè se il contenuto è insensibile alla distinzione tra le due regioni ambientali. La s-separabilità cattura l'idea che, relativamente a quella partizione dell'ambiente, il sistema si comporta come se il contenuto dipendesse solo dallo stato interno. Analogamente, la *e-separabilità* rispetto a una partizione $S = S_1 \cup S_2$ cattura l'idea che, in quella regione dello spazio degli stati, il contenuto dipende solo dall'ambiente.

La Terra Gemella putnamiana è ora descrivibile come la seguente situazione: se E_T è l'ambiente terrestre (contenente H_2O) ed E_TG è l'ambiente della Terra Gemella (contenente XYZ), e se s_0 è lo stato interno condiviso da Oscar e Oscar*, allora:

$$\mathscr{C}(s_0, \text{E_T}) \neq \mathscr{C}(s_0, \text{E_TG})$$

La funzione di contenuto non è s-separabile rispetto alla partizione {E_T, E_TG}: le due regioni ambientali producono contenuti diversi a parità di stato interno. Ma questo, nella prospettiva relazionalista, non dimostra che il contenuto sia "fuori dalla testa": dimostra semplicemente che la funzione $\mathscr{C}$ ha una derivata parziale rispetto ad e non nulla in quel punto. La conclusione interessante non è "il contenuto è esterno" ma "la funzione di contenuto è sensibile a questa particolare distinzione ambientale". La domanda successiva — ed è qui che il relazionalismo offre strumenti nuovi — è: *perché* è sensibile proprio a quella distinzione, e non ad altre?

La risposta putnamiana implicita è: perché H_2O e XYZ sono tipi naturali genuinamente distinti, e il riferimento di "water" è determinato causalmente dalla struttura del tipo naturale. Ma questa risposta, nel quadro relazionalista, deve essere completata: la derivata parziale $\partial\mathscr{C}/\partial e$ è non nulla rispetto alla distinzione H_2O/XYZ perché il sistema cognitivo in questione è *inserito causalmente* in un ambiente di tipo H_2O, e questa inserzione causale fa parte della struttura dello spazio $S \times E$ rilevante per quel sistema. In altri termini: la struttura di $S \times E$ non è il prodotto cartesiano di due insiemi arbitrari; è un prodotto *strutturato* dalla relazione causale che connette il tipo di sistema al tipo di ambiente. Si tratta di ciò che chiamo la *curvatura semantica* dello spazio $S \times E$: lo spazio prodotto non è piatto ma ha una geometria determinata dalla storia causale del sistema.

Questa idea richiede un'ulteriore formalizzazione. Introduco una metrica sullo spazio $S \times E$ che cattura la *prossimità semanticamente rilevante* tra coppie di stati-ambiente. Due punti (s_1, e_1) e (s_2, e_2) sono vicini semanticamente se producono contenuti simili:

$$d_sem((s_1, e_1), (s_2, e_2)) = d_P(\mathscr{C}(s_1, e_1), \mathscr{C}(s_2, e_2))$$

dove d_P è una metrica appropriata sullo spazio P delle proposizioni. La curvatura semantica è allora la curvatura della metrica d_sem rispetto alle coordinate "interne" (s) e "ambientali" (e) separate. Un'alta curvatura nella direzione e a stato interno fisso indica che piccole variazioni ambientali producono grandi variazioni di contenuto: il sistema è semanticamente sensibile all'ambiente in quella regione. Un'alta curvatura nella direzione s ad ambiente fisso indica che piccole variazioni interne producono grandi variazioni di contenuto.

Il progetto internalista foderiano equivale all'ipotesi che la curvatura semantica nella direzione e sia ovunque nulla o trascurabile: lo spazio S × E è semanticamente "piatto" rispetto alla coordinata ambientale. Il progetto esternalista putnamiano afferma che la curvatura nella direzione e è non nulla in regioni semanticamente decisive, come quella che distingue H_2O da XYZ. Il relazionalismo semantico afferma che la curvatura è strutturalmente bidirezionale, e che la forma geometrica dello spazio semantico — inclusa la distribuzione delle curvature — è ciò che occorre determinare empiricamente e teoricamente, senza precludere a priori la risposta.

XII. Il problema della circolarità ridisegnato: il bootstrap semantico

Una possibile obiezione al relazionalismo semantico è che esso sposti il problema senza risolverlo. Se il contenuto è definito sull'intero spazio S × E, e se la struttura di questo spazio dipende dalla relazione causale tra il sistema e l'ambiente, occorre ancora caratterizzare questa relazione causale — e qualsiasi caratterizzazione utilizzerà termini che presuppongono, implicitamente, una nozione di contenuto. La circolarità denunciata nell'internalismo e nell'esternalismo riappare, in forma raffinata, anche nel relazionalismo.

Rispondo che questa obiezione è corretta nella lettera ma equivoca nello spirito. La circolarità che affligge l'internalismo e l'esternalismo è *fondativa*: essi pretendono di individuare il confine del soggetto (o il ruolo dell'ambiente) *prima* di introdurre il contenuto, come base ontologica su cui il contenuto si costruisce. La circolarità che il relazionalismo ammette è *metodologica* e *coerenziale*: il relazionalismo non pretende di partire da una base ontologica pre-semantica, ma di caratterizzare una struttura reciprocamente vincolante in cui soggetto, ambiente e contenuto si co-determinano. Questo tipo di circolarità — che preferisco chiamare *bootstrap semantico* — non è un difetto ma una caratteristica strutturale di qualsiasi teoria dell'intenzionalità che prenda sul serio l'inscindibilità della mente e del mondo.

Il bootstrap semantico ha una struttura formale precisa. Sia $\mathscr{S}$ la teoria semantica, intesa come un insieme di asserzioni che connettono stati del sistema cognitivo,

proprietà dell'ambiente e contenuti proposizionali. $\mathscr{S}$ è internamente coerente se e solo se:

$\forall \langle s, e, p \rangle \in \mathscr{S}$: $\mathscr{C}(s, e) = p$ è compatibile con le proprietà strutturali di $\mathscr{C}$

La coerenza non richiede che esista un punto di partenza non semantico da cui tutto si deriva: richiede che l'insieme di asserzioni formi un sistema compatto, in cui ogni componente è giustificata dalla coerenza con le altre. Si tratta di una struttura epistemica analoga a quella della fisica teorica, dove i concetti fondamentali — spazio, tempo, massa, campo — si definiscono reciprocamente attraverso le equazioni della teoria, e non a partire da nozioni pre-teoriche indipendenti. Il fatto che non vi sia un "primo" concetto da cui tutti gli altri derivano non è un difetto della fisica: è una caratteristica della maturità teorica di un campo che ha abbandonato le fondazioni pre-teoriche in favore di una rete di vincoli mutuamente sostenentisi.

Il relazionalismo semantico aspira a un analogo stadio di maturità: non fondare la semantica su ontologie pre-date (la testa, il mondo) ma costruire una struttura in cui contenuto, sistema e ambiente sono articolati da vincoli reciproci la cui coerenza è la garanzia della loro adeguatezza.

Curvatura Semantica, Teleologia e lo Zombie Relazionale

XIII. Perché la teleosemantica non risolve il problema ma lo reintroduce

La teleosemantica — elaborata indipendentemente da Ruth Millikan e da Fred Dretske negli anni Ottanta e Novanta — è stata salutata come la risposta naturista al problema del contenuto: un modo per ancorare il contenuto mentale alla storia biologica del sistema cognitivo senza cadere nell'arbitrarietà soggettivistica dell'internalismo né nell'astrattezza dell'esternalismo nomologico putnamiano. L'intuizione fondamentale è che il contenuto di una rappresentazione sia determinato dalla sua *funzione propria* (*proper function*) nel senso di Millikan: la funzione che i predecessori causali della rappresentazione devono aver svolto perché i sistemi che la producono siano stati selezionati per produrla. Una rappresentazione ha contenuto C se e solo se la sua funzione propria è di corrispondere agli stati del mondo che istanziano C, e questa funzione propria è determinata dalla storia evolutiva del sistema rappresentazionale.

Il fascino della teleosemantica è che essa sembra fornire esattamente ciò che manca all'esternalismo putnamiano: un ancoraggio causale del contenuto che non dipende dalla competenza semantica dei parlanti né dalle convenzioni linguistiche della comunità, ma da fatti biologici oggettivi. La dipendenza del contenuto dall'ambiente

non è introdotta per via di esperimento mentale ma giustificata evolutivamente: i sistemi cognitivi si sono selezionati in un certo ambiente, e il loro contenuto è determinato da quella storia selettiva.

Ritengo tuttavia che la teleosemantica, nonostante la sua eleganza, reintroduca il problema che intende risolvere, e lo faccia a un livello di maggiore occultamento. Il problema si presenta sotto forma di ciò che chiamo *indeterminatezza del livello selettivo*.

La funzione propria di un sistema è determinata dalla storia selettiva. Ma la storia selettiva non individua univocamente la funzione: essa individua una classe di funzioni possibili, tutte compatibili con la stessa pressione selettiva. Un meccanismo che si è selezionato perché consente all'organismo di rilevare acqua potabile nell'ambiente si è selezionato per rilevare H_2O? O per rilevare il liquido incolore e insapore che disseta? O per rilevare ciò che gli altri organismi della stessa specie bevono? Queste tre funzioni sono coestensive nell'ambiente di adattamento originario — producono le medesime risposte comportamentali nelle medesime circostanze — ma divergono quando si vari l'ambiente. Non vi è, nella pressione selettiva stessa, un fatto che distingua quale dei tre descrizioni è la funzione propria corretta.

Millikan è consapevole di questa obiezione e risponde distinguendo la funzione propria *diretta* dalla funzione propria *derivata*, e introducendo la nozione di *condizioni normali* (Normal conditions) in cui la funzione si svolge correttamente. Ma queste condizioni normali sono a loro volta determinate dalla storia selettiva, e si ricade nello stesso problema: le condizioni normali dell'ambiente selettivo originario non individuano univocamente il contenuto nei casi controfattuali interessanti.

La difficoltà può essere formalizzata nel quadro relazionalista sviluppato nella sezione precedente. La teleosemantica propone, implicitamente, di determinare la funzione di contenuto $\mathscr{C}$ attraverso la storia evolutiva: il valore di $\mathscr{C}(s, e)$ per ambienti e non compresi nell'ambiente di adattamento è determinato dall'*estensione* della funzione selettiva agli ambienti non incontrati. Ma questa estensione non è univocamente determinata: vi sono molteplici modi di estendere la funzione $\mathscr{C}$ al di là del dominio di addestramento evolutivo, e la storia selettiva non discrimina tra essi. Formalmente: se $E_A \subset E$ è il sottoinsieme degli ambienti di adattamento, e se $\mathscr{C}$ è nota su $S \times E_A$, vi sono infinitely molte estensioni di $\mathscr{C}$ a $S \times E$ compatibili con i dati evolutivi. La teleosemantica non fornisce un criterio per scegliere tra queste estensioni.

Questo non è un problema tecnico riparabile con raffinamenti: è una difficoltà di principio. La storia causale — sia essa causal-nomologica come in Dretske o selettivo-biologica come in Millikan — determina il contenuto per i casi paradigmatici ma lascia strutturalmente indeterminato il contenuto nei casi controfattuali. E sono precisamente i casi controfattuali — la Terra Gemella, l'artrite alla coscia, lo zombie filosofico — che costituiscono il banco di prova di qualsiasi teoria del contenuto.

Nel quadro relazionalista, la teleosemantica equivale all'ipotesi che $\mathscr{C}$ sia *rigidamente estendibile* a partire dal suo valore su $S \times E_A$ — che esista un'unica estensione "naturale" compatibile con la pressione selettiva. Ma la struttura geometrica dello spazio semantico (la curvatura di $\mathscr{C}$ in $S \times E$) mostra che non vi è alcuna garanzia di

unicità: la curvatura locale su E_A non determina la curvatura globale su E. La teleosemantica postula implicitamente una "piattezza" semantica dello spazio E che non è affatto garantita, e che si rivela falsa nei casi controfattuali interessanti.

XIV. Lo zombie relazionale e il limite interno del relazionalismo

La critica alla teleosemantica riporta il relazionalismo semantico davanti al proprio limite interno — quello che considero la sua difficoltà più seria e che intendo affrontare senza minimizzarla. Il limite si manifesta attraverso un esperimento mentale che chiamo *lo zombie relazionale*.

Uno zombie relazionale è un sistema che, per ogni coppia $\langle s, e \rangle \in S \times E$, produce la stessa funzione di contenuto $\mathscr{C}$ di un sistema cognitivo cosciente, ma in assenza di qualsiasi comprensione o intenzionalità genuina. Se la funzione $\mathscr{C}$ è definita completamente in termini di mappature input-output — come sembra quando si dice che il contenuto è una funzione dallo spazio $S \times E$ allo spazio P — allora uno zombie relazionale è concettualmente possibile: un sistema che computa $\mathscr{C}$ correttamente senza "capire" nulla. Questo è, naturalmente, una versione del problema originale di Searle con la Stanza Cinese, ma nella formulazione relazionalista acquista una dimensione specifica.

La risposta che il relazionalismo deve dare non può essere semplicemente funzionalista — non può dire che lo zombie relazionale è impossibile perché avere la funzione $\mathscr{C}$ giusta *è* avere il contenuto. Questa risposta richiuderebbe il relazionalismo nel funzionalismo standard, che ha le proprie difficoltà ben note. Il relazionalismo deve invece chiarire in che senso la struttura differenziale di $\mathscr{C}$ — la forma geometrica dello spazio semantico — è una proprietà non del sistema isolato né dell'ambiente isolato, ma della *relazione* tra sistema e ambiente. E questa relazione non è una relazione astratta tra insiemi: è una relazione causale, dinamica, storicamente situata.

Propongo di introdurre a questo scopo la distinzione tra due livelli della funzione di contenuto: il livello *estensionale* e il livello *intensionale-relazionale*.

Al livello estensionale, $\mathscr{C}$ è semplicemente una funzione matematica: una mappa da $S \times E$ a P. Questo livello è quello che uno zombie relazionale potrebbe in principio replicare. Al livello intensionale-relazionale, $\mathscr{C}$ non è soltanto una mappa ma *il modo in cui* il sistema si relaziona all'ambiente — la struttura del legame causale e normativo che connette stati interni a situazioni ambientali. Questo livello non è replicabile da uno zombie perché non è definito da un insieme di coppie ⟨input, output⟩ ma dalla *natura* della relazione causale che le genera.

La distinzione può essere precisata come segue. Diciamo che due sistemi A e B sono *estensionalmente equivalenti* rispetto a $\mathscr{C}$ se:

$\forall (s, e) \in S \times E : \mathscr{C}_A(s, e) = \mathscr{C}_B(s, e)$

Diciamo che sono *intensionalmente equivalenti* se, in aggiunta, la struttura causale che genera le due funzioni è del medesimo tipo — se le due funzioni sono prodotte da meccanismi dello stesso genere causale. Lo zombie relazionale è estensionalmente equivalente al sistema cosciente ma non intensionalmente equivalente: produce la stessa mappa ma attraverso un meccanismo causalmente diverso.

Il relazionalismo semantico sostiene che ciò che conta per il contenuto genuino è l'equivalenza intensionale, non quella estensionale. Ma questa affermazione solleva immediatamente la domanda: cosa rende la struttura causale rilevante? E qui il relazionalismo si trova al confine con il problema difficile della coscienza: perché certe strutture causali sono associate a contenuto genuino e altre no? Non ho una risposta completa a questa domanda, e sarebbe disonesto fingere di averla. Ciò che il relazionalismo offre è una ridefinizione del problema: non "come determina l'ambiente il contenuto" ma "quale tipo di relazione causale tra sistema e ambiente costituisce contenuto genuino piuttosto che mera mappatura funzionale". La risposta a questa domanda richiede una teoria della causalità sufficientemente ricca da distinguere relazioni causali di tipi diversi — e questa teoria non è ancora disponibile in forma soddisfacente.

XV. La curvatura semantica come invariante: un'ipotesi positiva

Voglio concludere questa parte con un'ipotesi positiva che, pur ammettendo il limite interno appena discusso, offre un contributo genuino al problema. L'ipotesi riguarda la *covarianza* della struttura semantica sotto trasformazioni del sistema e dell'ambiente.

Nel quadro relazionalista, il contenuto di uno stato mentale non è un oggetto fisso ma una struttura locale: è il differenziale di $\mathscr{C}$ in un punto $(s, e) \in S \times E$. Ma vi sono aspetti di questa struttura locale che rimangono invarianti sotto certe trasformazioni. Propongo di chiamare *invarianti semantici* quelle proprietà di $\mathscr{C}$ che non cambiano quando si compono trasformazioni compatibili con la struttura causale del sistema.

L'idea si chiarisce con un esempio. Supponiamo che Oscar si trasferisca sulla Terra Gemella. Con il passare del tempo, le sue interazioni causali con XYZ rideterminerano parzialmente la struttura della sua funzione $\mathscr{C}$: ciò che egli chiama "water" inizierà a riferirsi ad XYZ piuttosto che ad H_2O, almeno per i casi localmente accessibili. Questo processo graduale di rideterminazione del contenuto corrisponde, nel quadro relazionalista, a una *deformazione* della funzione $\mathscr{C}$ nella regione $S \times E_TG$. La domanda interessante non è "Oscar pensa a H_2O o a XYZ?" — domanda che presuppone un contenuto fisso — ma "qual è la velocità e la struttura della deformazione di $\mathscr{C}$ man mano che Oscar interagisce con XYZ?".

Gli invarianti semantici sono quelle proprietà di $\mathscr{C}$ che rimangono stabili durante questa deformazione. Una proprietà plausibilmente invariante è la *topologia locale* dello spazio semantico: la struttura di quali contenuti sono "vicini" tra loro, quale curvatura ha lo spazio P nella regione raggiunta dal punto $\mathscr{C}$(s, e). Questa topologia è determinata dalla struttura della mente di Oscar — dalle sue capacità percettive, dalle sue disposizioni inferenziali — e non dalla struttura molecolare dell'ambiente. In questo senso limitato, vi è un contributo "interno" alla struttura del contenuto che sopravvive alla variazione ambientale: non il contenuto determinato, ma la *forma* dello spazio di contenuti accessibili.

Questo è ciò che il relazionalismo semantico può concedere all'internalismo senza diventarlo: non che il contenuto sia stretto, ma che vi è uno *strato strutturale invariante* della funzione di contenuto — la sua topologia locale — che dipende dal sistema indipendentemente dall'ambiente specifico. Questo strato invariante è ciò che rende possibile il riconoscimento — il fatto che Oscar, giunto sulla Terra Gemella, trovi XYZ *simile* all'acqua, e non radicalmente diverso come trova una roccia o un'idea astratta. La similarità è determinata dalla topologia dello spazio semantico locale, che è una proprietà del sistema cognitivo; il contenuto determinato è determinato dalla posizione in S × E, che è una proprietà relazionale.

Il relazionalismo semantico è dunque una posizione *stratificata*: riconosce uno strato strutturale (la topologia dello spazio semantico) che dipende prevalentemente dal sistema, e uno strato determinativo (il contenuto effettivo in ogni situazione) che dipende irriducibilmente dalla relazione sistema-ambiente. Né puro internalismo né puro esternalismo: una geometria semantica a due livelli, in cui il primo fissa la forma dello spazio e il secondo localizza il contenuto in quello spazio.

Leggi Psicologiche sotto il Relazionalismo, Indicali e Riferimento Singolare

---

XVI. Il destino delle leggi psicologiche in un'ontologia relazionale

La preoccupazione foderiana per le leggi psicologiche non è capricciosa: è l'espressione di un'esigenza metodologica genuina. Se la psicologia cognitiva deve essere una scienza con potere esplicativo e predittivo, le sue generalizzazioni devono avere una portata che trascenda il caso singolo. Una legge del tipo "se un agente crede che P e desidera che Q, e ritiene che fare A sia sufficiente perché Q, allora, ceteris paribus, farà A" deve essere applicabile a qualsiasi agente che si trovi nella condizione descritta, indipendentemente da dove o quando vive, da quale ambiente lo circonda, da quale storia causale lega i suoi stati interni al mondo. È questa universalità applicativa — questa insensibilità alle variazioni ambientali — che Fodor intende

proteggere esigendo che le leggi psicologiche siano formulate in termini di contenuto stretto.

Il relazionalismo semantico sembra minacciare questa universalità. Se il contenuto è irriducibilmente relazionale — se $\mathscr{C}(s, e)$ dipende dall'ambiente e — allora due agenti in ambienti diversi, pur essendo intrinsecamente identici, hanno contenuti diversi, e le leggi psicologiche formulate in termini di contenuti non si applicheranno uniformemente ad entrambi. La legge pratica sopra enunciata, se "P", "Q" e "fare A" denotano contenuti larghi, varrà per Oscar nell'ambiente terrestre ma non necessariamente per Oscar* nell'ambiente della Terra Gemella, poiché i contenuti di Oscar* sono parzialmente diversi.

Questa obiezione, per quanto pressante, si basa su un equivoco sottile riguardo alla struttura delle leggi nelle scienze speciali. Le leggi delle scienze speciali non sono, in genere, leggi sugli oggetti individuati da proprietà di livello inferiore: sono leggi su *tipi* di oggetti individuati a un certo livello di descrizione, e la loro universalità è relativa a quel livello. La legge di Gresham — la moneta cattiva scaccia quella buona — non è una legge sulle proprietà fisiche dei pezzi metallici: è una legge su un tipo di oggetto (la moneta) individuato dalle sue proprietà funzionali ed istituzionali, e vale per qualsiasi sistema che abbia quelle proprietà funzionali, indipendentemente dalla sua realizzazione fisica. La questione è se le leggi psicologiche abbiano la medesima struttura: leggi su tipi di stati mentali individuati a un certo livello di descrizione, con universalità relativa a quel livello.

Il relazionalismo semantico permette di riformulare questa questione con precisione. Nel quadro relazionalista, i tipi di stati mentali rilevanti per le leggi psicologiche non sono individuati né dai soli stati interni né dai soli contenuti larghi: sono individuati dalla *struttura locale della funzione di contenuto* — dalla curvatura semantica in un intorno del punto $(s, e) \in S \times E$. Due stati mentali appartengono allo stesso tipo psicologico se e solo se la struttura locale di $\mathscr{C}$ intorno ai rispettivi punti è dello stesso tipo — se, cioè, hanno la stessa forma differenziale locale.

Formalmente: due coppie (s_1, e_1) e (s_2, e_2) sono psicologicamente equivalenti, e denotano istanze dello stesso tipo mentale, se esiste un diffeomorfismo locale $\varphi : U_1 \to U_2$ (dove U_1 e U_2 sono intorni aperti dei due punti in $S \times E$) tale che:

$$\mathscr{C} \circ \varphi = \mathscr{C} \quad \text{su } U_1$$

— cioè se la restrizione di $\mathscr{C}$ a U_1 è diffeomorficamente equivalente alla restrizione a U_2. L'equivalenza diffeomorfica locale cattura l'idea che i due stati abbiano la stessa struttura semantica locale, ossia che il "modo" in cui essi variano con variazioni di s e di e sia strutturalmente il medesimo, anche se i contenuti effettivi che producono sono diversi.

Le leggi psicologiche, su questa concezione, sono leggi su *classi di equivalenza diffeomorfica* della funzione di contenuto: esse affermano che coppie appartenenti alla stessa classe di equivalenza si comportano uniformemente rispetto a certe transizioni.

Questo ridefinisce l'universalità delle leggi psicologiche: non universalità rispetto agli stati interni (come vorrebbe Fodor) né rispetto ai contenuti larghi (come implicherebbe la psicologia del senso comune ingenua), ma universalità rispetto alla struttura differenziale locale della funzione di contenuto. Si tratta di un'universalità *geometrica*, non sintattica né semantica nel senso tradizionale.

Questa proposta ha una conseguenza di notevole interesse: essa implica che la psicologia cognitiva sia, nella sua struttura profonda, una disciplina la cui oggetto formale è la geometria dello spazio semantico — la forma dello spazio $S \times E$ equipaggiato con la funzione di contenuto $\mathscr{C}$. Non è una psicologia della sintassi (Fodor) né una psicologia del contenuto (la folk psychology), ma una psicologia della *struttura differenziale del contenuto*. Il programma di ricerca che ne segue non è la ricerca di stati interni universalmente condivisi, né la descrizione di relazioni soggetto-mondo caso per caso: è la classificazione delle strutture geometriche locali dello spazio semantico, e l'identificazione delle leggi di trasformazione che mappano strutture in strutture.

XVII. Indicali, centri e la peculiarità del contenuto di primo grado

Tra tutti i fenomeni semantici, gli indicali — i termini il cui riferimento dipende sistematicamente dal contesto di proferenza — sono quelli che sfidano più direttamente qualsiasi teoria del contenuto, sia internalista che esternalista. Un indicale come "io", "qui", "adesso", "questo" non ha un contenuto fisso: il suo riferimento varia con il parlante, il luogo, il momento, l'oggetto puntato. Qualsiasi teoria del contenuto deve rendere conto di questo fenomeno, e il modo in cui lo fa rivela assunzioni strutturali profonde.

La teoria standard degli indicali, elaborata da David Kaplan nel suo *Demonstratives* (1977, pubblicato come articolo nel 1989), distingue due livelli di significato: il *carattere* di un'espressione indicale — una funzione dal contesto al contenuto — e il *contenuto* dell'espressione relativa a un contesto — la proposizione espressa, o il contributo proposizionale del termine. Il carattere di "io" è la regola semantica che assegna ad ogni contesto di proferimento il parlante di quel contesto come contributo al contenuto. Il contenuto di "io" come proferito da Oscar in un certo contesto è Oscar stesso — un individuo, non una descrizione.

La struttura kaplaniana è elegante e potente, ma pone un problema specifico per il dibattito stretto/largo. I contenuti kaplaniani sono *larghi* per costruzione: il contenuto di "io" proferito da Oscar è Oscar, non una proprietà interna di Oscar. Ma il carattere — la regola che mappa contesti in contenuti — sembra essere qualcosa di più vicino allo stretto: è una proprietà dell'espressione indipendente dal contesto, e due parlanti che condividano la competenza semantica per "io" condividono il medesimo carattere. Il carattere è, in un senso preciso, l'analogo kaplaniano del contenuto stretto: ciò che

sopravviene sulla competenza linguistica del parlante indipendentemente dal suo ambiente.

Ma questa analogia è ingannevole, e l'inganno si manifesta nel modo in cui il contenuto di primo grado — il contenuto dei pensieri indicali, dei pensieri de se e de nunc — si inserisce nella struttura kaplaniana. Il problema è quello che John Perry ha chiamato l'*essential indexical*: vi sono credenze la cui forza esplicativa dipende ineliminabilmente dalla loro struttura indicale, e queste credenze non possono essere parafrasate in termini non indicali senza perdere il loro potere causale. Quando Perry racconta di aver inseguito il carrello del supermercato che lasciava una scia di zucchero senza rendersi conto di essere egli stesso la persona che perdeva lo zucchero, il momento cruciale della comprensione è la formazione della credenza "sono *io* quello che perde lo zucchero" — una credenza de se che non è riducibile a nessuna credenza de re o descrittiva.

Nel quadro relazionalista, i pensieri indicali hanno una struttura della funzione di contenuto $\mathscr{C}$ che differisce strutturalmente da quella dei pensieri non indicali. Per un pensiero indicale, la dipendenza da e — la derivata parziale $\partial\mathscr{C}/\partial$e — ha una forma speciale: non è la dipendenza dal tipo di ambiente (come nel caso dell'acqua e di H_2O) ma la dipendenza dalle *coordinate* del soggetto nell'ambiente — dalla sua posizione spaziotemporale, dalla sua identità come nodo nella rete causale. Il contenuto di "io" non dipende da cosa c'è nel mondo ma da *dove* e *chi* è il sistema cognitivo nel mondo.

Questa distinzione può essere formalizzata introducendo una decomposizione dello spazio E degli ambienti in due componenti: una componente *tipologica* E_T — che specifica le proprietà generali dell'ambiente, come la composizione molecolare dell'acqua — e una componente *posizionale* E_P — che specifica la posizione e l'identità del sistema cognitivo nell'ambiente. Un pensiero non indicale ha una funzione di contenuto sensibile principalmente ad E_T: $\mathscr{C}$(s, e) dipende da e tramite la componente tipologica. Un pensiero indicale ha una funzione di contenuto sensibile principalmente ad E_P: $\mathscr{C}$(s, e) dipende da e tramite la coordinata posizionale del soggetto.

Per pensieri non indicali: $\partial\mathscr{C}/\partial$e_T $\neq 0$, $\partial\mathscr{C}/\partial$e_P ≈ 0
Per pensieri indicali: $\partial\mathscr{C}/\partial$e_T ≈ 0, $\partial\mathscr{C}/\partial$e_P $\neq 0$

Il carattere kaplaniano corrisponde, in questa formalizzazione, alla struttura della dipendenza da E_P: la regola che mappa posizioni nell'ambiente a contenuti. Il contenuto kaplaniano corrisponde al valore di $\mathscr{C}$(s, e_P) per una posizione specifica. La distinzione carattere/contenuto è dunque, nel quadro relazionalista, una distinzione tra due livelli della funzione $\mathscr{C}$: il livello della struttura della dipendenza posizionale e il livello del valore specifico in una posizione data.

XVIII. Il riferimento singolare e il problema dell'ancoraggio

Il riferimento singolare — la capacità di pensieri e termini di agganciarsi a un individuo specifico piuttosto che a un tipo o a una proprietà — ha un posto privilegiato nel dibattito stretto/largo perché sembra resistere tanto all'internalismo quanto all'esternalismo semplice. Il problema dell'ancoraggio è il seguente: per quale ragione un certo stato mentale si riferisce a *questo* individuo piuttosto che a un altro intrinsecamente simile? Qualsiasi risposta deve fare appello a qualcosa che distingue i due individui — ma tale distinzione non può essere una proprietà interna del sistema cognitivo, perché due individui distinti possono essere intrinsecamente identici, e i sistemi cognitivi che li rappresentano possono essere a loro volta intrinsecamente identici.

La risposta causalista — il riferimento è determinato dalla catena causale che connette il termine all'individuo — è la più naturale, ma incontra il problema della *contaminazione causale*. Le catene causali raramente sono lineari e univoche: il termine "Aristotele" nella mente di un parlante contemporaneo è connesso causalmente non solo ad Aristotele ma a migliaia di usi intermedi del termine, a libri, a enciclopedie, a errori e a correzioni. Come si individua, all'interno di questa rete causale complessa, la catena rilevante che connette il termine al suo referente originario?

Nel quadro relazionalista, il problema dell'ancoraggio si presenta come il problema di determinare quale struttura della funzione di contenuto $\mathscr{C}$ corrisponde al riferimento singolare. Propongo che il riferimento singolare non sia una proprietà *puntuale* di $\mathscr{C}$ — non dipenda da un valore specifico di $\mathscr{C}(s, e)$ — ma una proprietà *globale* della struttura di $\mathscr{C}$ in una regione di $S \times E$. Più precisamente, un sistema cognitivo si riferisce singolarmente a un individuo x se e solo se esiste una regione $R \subseteq S \times E$ tale che:

(i) $\forall (s, e) \in R : x \in \text{mondo}(e) \rightarrow \mathscr{C}(s, e)$ include x come costituente
(ii) La regione R è connessa e causalmente coerente — è raggiungibile da una catena di transizioni causali a partire da un contatto percettivo o dimostrativo originario con x
(iii) La funzione $\mathscr{C}$ è stabile su R rispetto a perturbazioni di s: piccole variazioni dello stato interno non rimuovono x dal contenuto

La condizione (i) cattura il requisito standard del riferimento: in tutti i contesti rilevanti, il referente è il costituente del contenuto. La condizione (ii) incorpora la dimensione causale: la regione R non è arbitraria ma è strutturata dalla storia causale del sistema, dall'atto di ancoraggio originario che ha stabilito la connessione con x. La condizione (iii) aggiunge un requisito di robustezza: il riferimento singolare non è un accidente di un particolare stato interno ma una proprietà stabile della funzione di contenuto in quella regione — resistente a variazioni cognitive ordinarie.

Questa caratterizzazione del riferimento singolare ha una conseguenza rilevante per il dibattito stretto/largo: essa mostra che il riferimento singolare è intrinsecamente *storico* nel senso in cui non può essere catturato da nessuna istantanea dello stato interno o

dell'ambiente. La condizione (ii) richiede che la regione R sia raggiungibile da un atto di ancoraggio originario: questa raggiungibilità è una proprietà della struttura causale dell'intera storia del sistema cognitivo, non di un suo stato attuale. Il riferimento singolare è, per così dire, una proprietà della *traiettoria* del sistema nello spazio $S \times E$, non di un suo punto.

Questo ha ripercussioni sull'identità dei pensieri nel tempo. Se mi ricordo di Platone — se ho un pensiero de re su Platone — la continuità del mio riferimento a Platone attraverso le successive occasioni di pensiero non è garantita dalla sola struttura interna dei miei stati: è garantita dalla continuità causale della mia storia cognitiva, dalla traiettoria che, risalendo di catena causale in catena causale, conduce fino agli esemplari antichi del nome "Platone". Spezzare questa continuità causale — per esempio, nel caso di una grave amnesia che cancelli la storia semantica del soggetto preservando le capacità inferenziali — potrebbe spezzare il riferimento pur lasciando intatta la struttura interna. L'identità del pensiero singolare è una proprietà della traiettoria, non del punto.

XIX. Conseguenze per la normatività semantica: chi corregge chi?

Una delle dimensioni meno esplorate del dibattito stretto/largo riguarda la *normatività semantica*: il fatto che gli stati mentali e le espressioni linguistiche possano essere corretti o scorretti, appropriati o inappropriati, precisi o imprecisi nella loro applicazione al mondo. Quando Tizio usa "artrite" per riferirsi al dolore muscolare alla coscia, il suo uso è semanticamente *scorretto* — e questa scorrettezza non è determinata da Tizio stesso ma dalla norma sociale che governa il termine. La normatività semantica sembra dunque avere una struttura intrinsecamente esternalista: i criteri di correttezza sono in qualche misura esterni al soggetto.

Il relazionalismo semantico permette di articolare questa normatività in modo più fine di quanto consentano le alternative disponibili. La normatività non è una proprietà aggiunta al contenuto dall'esterno: è una struttura interna alla funzione di contenuto $\mathscr{C}$. Più precisamente, la normatività semantica corrisponde alla nozione di *sezione* della funzione di contenuto: una sezione è una scelta, per ogni stato interno s, di un ambiente $e^*(s)$ "canonico" tale che $\mathscr{C}(s, e^*(s))$ rappresenta il contenuto *corretto* di s. La norma semantica è la sezione — la scelta dell'ambiente canonico rispetto al quale gli stati mentali sono valutati come corretti o scorretti.

Formalmente, una sezione $\sigma : S \to E$ è una mappa che assegna ad ogni stato interno uno stato dell'ambiente "di default", e la funzione di contenuto normativo è:

$$\mathscr{C}_\sigma : S \to P$$
$$\mathscr{C}_\sigma(s) = \mathscr{C}(s, \sigma(s))$$

Il contenuto normativo — il contenuto valutato rispetto alla sezione canonica — è una funzione *solo* su S: è determinato dallo stato interno più l'ambiente canonico. L'internalismo è il caso limite in cui la sezione σ è la stessa per tutti gli ambienti possibili — tutti gli ambienti sono equivalenti rispetto alla norma semantica. L'esternalismo di Burge corrisponde al caso in cui la sezione σ è determinata dalle convenzioni della comunità linguistica: l'ambiente canonico non è l'ambiente reale del soggetto ma l'ambiente normativo — il mondo come la comunità esperta lo classifica.

La domanda "chi corregge chi?" — se è la comunità a determinare il contenuto del soggetto o se è il soggetto a dover conformare il proprio uso alla realtà — non ha nel relazionalismo una risposta univoca. Vi sono diversi tipi di sezioni canoniche, e la normatività semantica è stratificata: vi è una normatività rispetto alle convenzioni linguistiche (la sezione burghiana), una rispetto alla struttura causale del mondo (la sezione putnamiana), una rispetto alla storia evolutiva del sistema (la sezione millikaniana). Queste sezioni possono divergere — e quando divergono, producono tensioni normative che il parlante ordinario sperimenta come incertezza semantica, come il senso di usare un termine in modo forse impreciso senza sapere come renderlo più preciso.

Il caso di Tizio e dell'artrite è, in questa luce, un caso di *conflitto tra sezioni*: la sezione determinata dall'uso individuale di Tizio diverge dalla sezione determinata dalla norma medica istituzionalizzata. Il contenuto del pensiero di Tizio non è né quello che Tizio crede di pensare né quello che il medico attribuirebbe: è il risultato della tensione tra sezioni diverse, una tensione che si risolve pragmaticamente in funzione del contesto comunicativo e degli scopi esplicativi di chi attribuisce il pensiero. Non vi è un fatto semantico puro e indipendente che determini il contenuto vero di Tizio: vi sono diverse funzioni di contenuto valide rispetto a diverse sezioni canoniche, e la scelta tra esse non è semantica ma pragmatica e teorica.

Questa conclusione può sembrare relativista, ma non lo è nel senso deteriore. Non afferma che qualsiasi contenuto vada bene: afferma che la normatività semantica è pluridimensionale, che vi sono più dimensioni lungo le quali uno stato mentale può essere valutato, e che la scelta della dimensione rilevante dipende dallo scopo per cui si attribuisce il contenuto. Per la psicologia cognitiva — che vuole spiegare il comportamento — la sezione rilevante potrebbe essere quella dell'ambiente prossimo del soggetto. Per la semantica formale — che vuole rendere conto delle condizioni di verità — la sezione rilevante è quella della struttura del mondo. Per la filosofia del linguaggio sociale — che studia la comunicazione nella comunità — la sezione rilevante è quella delle norme condivise. Il relazionalismo semantico non elimina la normatività; la moltiplica e la situa.

Cognizione Distribuita, Identità Diacronica del Contenuto e Conclusioni

---

XX. Il contenuto nelle architetture distribuite: un banco di prova radicale

Il relazionalismo semantico trova il suo banco di prova più esigente nelle architetture cognitive distribuite — sia quelle biologiche (i sistemi cognitivi collettivi, le colonie sociali, le reti di organismi in simbiosi cognitiva) sia quelle artificiali (i sistemi di apprendimento profondo, le reti neurali artificiali addestrate su corpora linguistici, le architetture multi-agente). Questi sistemi pongono la domanda del contenuto in una forma che le teorie tradizionali non sono equipaggiate per affrontare: se il sistema cognitivo non è un singolo individuo ma una rete di nodi interagenti, dov'è localizzato il contenuto? Come si assegna la funzione di contenuto $\mathscr{C}$ a un sistema che non ha un singolo spazio degli stati interni S ma una molteplicità di spazi parziali interconnessi?

La risposta internalista si inceppa immediatamente: non vi è una "testa" in cui il contenuto possa essere localizzato, né una singola architettura interna rispetto alla quale la sopravvenienza possa essere definita. La risposta esternalista standard soffre di un problema simmetrico: l'ambiente di un sistema distribuito è non uniforme, ogni nodo interagisce con regioni diverse dell'ambiente, e non vi è una relazione soggetto-mondo unica che determini il contenuto dell'intero sistema.

Il relazionalismo semantico affronta questa situazione estendendo il formalismo in modo naturale. Per un sistema distribuito con nodi N_1, ..., N , lo spazio degli stati interni è il prodotto:

$$S_dist = S_1 \times S_2 \times ... \times S$$

dove S_i è lo spazio degli stati del nodo N_i. Lo spazio degli ambienti è egualmente strutturato: ogni nodo ha un ambiente parziale E_i, e l'ambiente globale è una struttura che comprende non solo gli ambienti locali ma le relazioni tra essi:

$$E_dist = E_1 \times E_2 \times ... \times E \times R(E_1,...,E)$$

dove $R(E_1,...,E)$ cattura le relazioni di coerenza tra gli ambienti locali — il fatto, cruciale, che gli ambienti locali dei diversi nodi siano parti di un ambiente globale condiviso e che le rappresentazioni di diversi nodi debbano, in qualche senso, coerire.

La funzione di contenuto del sistema distribuito è:

$$\mathscr{C}_dist : S_dist \times E_dist \rightarrow P_dist$$

dove P_dist è uno spazio di proposizioni che può essere più ricco di quello accessibile a qualsiasi singolo nodo. La struttura semantica del sistema distribuito non è riducibile alla somma delle strutture semantiche dei nodi individuali: vi sono contenuti che emergono dalla relazione tra nodi — contenuti *inter-nodali* — che nessun nodo porta

individualmente. L'informazione semantica, in questo senso, è genuinamente distribuita: non localizzabile, non riducibile a componenti.

Questo ha una conseguenza precisa per il dibattito stretto/largo: in un sistema distribuito, la distinzione non ha nemmeno un referente chiaro. Che cosa sarebbe il "contenuto stretto" di un sistema distribuito? Il contenuto che sopravviene su S_dist ignorando E_dist? Ma la struttura di S_dist è già, per costruzione, parzialmente determinata dalla struttura di E_dist, perché le architetture di connessione tra nodi rispecchiano le strutture causali dell'ambiente. La distinzione stretto/largo si dissolve non per assenza di distinzione ma per collasso dei due poli nell'unica struttura relazionale S_dist × E_dist.

Le reti neurali artificiali offrono qui un caso di studio particolarmente istruttivo, non come modelli della mente biologica ma come *esperimenti concettuali realizzati*. Una rete addestrata su un corpus linguistico sviluppa una struttura interna — pesi, attivazioni, geometria dello spazio delle rappresentazioni — che è interamente determinata dall'ambiente di addestramento. Non vi è separazione netta tra "struttura interna" e "influenza ambientale": la struttura interna *è* la sedimentazione dell'ambiente, la sua compressione in parametri. La funzione di contenuto di tali sistemi è dunque un caso estremo in cui $\partial \mathscr{C}/\partial e$ ha lasciato un'impronta permanente su S attraverso il processo di addestramento: l'ambiente ha *plasmato* lo spazio degli stati interni, non solo *influenzato* i valori di $\mathscr{C}$ dato S.

Questo suggerisce una generalizzazione del formalismo relazionalista che cattura la dimensione diacronica del contenuto. Lo spazio S degli stati interni non è statico: si deforma nel tempo per effetto delle interazioni con l'ambiente. La funzione di contenuto $\mathscr{C}$ non è definita su un prodotto fisso S × E ma su un prodotto *dinamico*, dove S(t) evolve in funzione della storia delle interazioni con E. Il contenuto di un sistema cognitivo maturo non è semplicemente la valutazione di $\mathscr{C}$ al tempo presente: è il risultato di un'intera traiettoria nel prodotto S(t) × E, una traiettoria che ha deformato S progressivamente rendendolo sensibile a certi aspetti di E e insensibile ad altri.

XXI. Identità diacronica del contenuto: il problema del cambiamento semantico

Il problema dell'identità diacronica del contenuto — come i pensieri mantengono o perdono la loro identità semantica nel corso del tempo — è uno dei più trascurati nella letteratura, eppure è di importanza fondamentale sia per la psicologia del sé sia per la comprensione del cambiamento concettuale nella scienza.

Consideriamo un caso semplice. Oscar, da bambino, acquisisce il concetto di "metallo" basandosi principalmente su caratteristiche percettive: lucentezza, durezza, conducibilità termica al tatto. Da adulto, dopo studi di chimica, il suo concetto di

"metallo" si arricchisce di struttura teorica: band theory della conduzione elettronica, struttura a reticolo cristallino, proprietà di ionizzazione. I due concetti — quello infantile e quello adulto — condividono lo stesso termine ma hanno strutture semantiche differenti. Sono lo stesso contenuto? O sono contenuti diversi che condividono un'etichetta?

La risposta internalista tende a dire che sono contenuti diversi, perché gli stati interni che li veicolano sono diversi. La risposta esternalista tende a dire che sono lo stesso contenuto, perché entrambi si riferiscono allo stesso tipo naturale (i metalli, individuati chimicamente) anche se con gradi diversi di conoscenza. Entrambe le risposte catturano qualcosa di vero ma mancano la struttura del fenomeno.

Nel quadro relazionalista, l'identità diacronica del contenuto corrisponde alla *continuità della traiettoria* nel prodotto $S(t) \times E$. Il contenuto adulto di Oscar è lo stesso di quello infantile, in un senso preciso, se e solo se la traiettoria $S(t)$ è *connessa* — se vi è un percorso continuo da S_bambino a S_adulto che non attraversa regioni di discontinuità semantica radicale. La continuità della traiettoria non richiede identità degli stati: richiede che ogni passo della trasformazione preservi sufficiente coerenza semantica locale. È l'analogo, nella semantica, della continuità dell'identità personale: non identità di substrato, ma continuità di struttura attraverso la trasformazione.

La nozione di continuità della traiettoria permette di distinguere il cambiamento semantico ordinario — che preserva l'identità del contenuto pur arricchendone la struttura — dalla *rottura semantica* — in cui la traiettoria attraversa una discontinuità tale da produrre contenuti genuinamente nuovi. Le rivoluzioni scientifiche kuhniane sono, in questa luce, casi di rottura semantica: quando la chimica post-lavoisieriana sostituisce la teoria del flogisto, non si tratta di arricchimento dello stesso contenuto ma di sostituzione, di discontinuità nella traiettoria semantica. I chimici post-rivoluzionari non pensano la stessa cosa dei predecessori con maggiore precisione: pensano cose strutturalmente diverse, anche quando usano i medesimi termini.

Questo ha implicazioni dirette per il problema dell'incommensurabilità kuhniana. Se l'incommensurabilità è una discontinuità nella traiettoria semantica — una rottura nella connessione della funzione di contenuto attraverso il cambiamento teorico — allora non è un difetto epistemico né un relativismo irrazionalista: è una caratteristica strutturale del cambiamento concettuale radicale, derivante dalla geometria dello spazio semantico. La connessione, o disconnessione, tra paradigmi scientifici è una proprietà della curvatura dello spazio semantico nella regione di transizione — un fatto strutturale che la storia della scienza può rilevare ma che la sola logica non può decidere.

XXII. Conclusioni: che cosa rimane del problema stretto/largo

Il percorso argomentativo di questa monografia ha condotto a una posizione che, nella sua forma compiuta, non è facilmente classificabile entro il dibattito tradizionale tra internalismo ed esternalismo. Conviene riepilogare le tesi principali con la precisione che meritano, distinguendo ciò che è stato dimostrato, ciò che è stato proposto come ipotesi e ciò che rimane aperto.

È stato dimostrato — e ritengo la dimostrazione cogente — che la distinzione stretto/largo nella sua formulazione standard è circolare: presuppone un confine soggetto/ambiente ontologicamente indipendente dal contenuto, ma tale confine è in realtà co-determinato con il contenuto. Questa circolarità non è un difetto tecnico corregibile: è il segnale di un'impostazione ontologica inadeguata, quella che ho chiamato dualismo di dominio.

È stato proposto come struttura teorica alternativa il relazionalismo semantico, secondo cui il contenuto è definito sull'intero spazio prodotto $S \times E$ come una funzione differenziale $\mathscr{C}$ la cui struttura locale — curvatura, separabilità, invarianti — cattura le proprietà semanticamente rilevanti degli stati mentali. Questa struttura permette di riformulare l'internalismo e l'esternalismo come condizioni limite su $\mathscr{C}$, di ridefinire le leggi psicologiche come leggi su classi di equivalenza diffeomorfica della funzione di contenuto, di caratterizzare il riferimento singolare come proprietà della traiettoria nello spazio $S \times E$, e di stratificare la normatività semantica in termini di sezioni canoniche diverse.

Rimane aperto il problema della distinzione tra equivalenza estensionale e intensionale della funzione di contenuto — il problema dello zombie relazionale — che si rivela essere, nella sua profondità, una manifestazione del problema difficile della coscienza in veste semantica. Il relazionalismo semantico non risolve questo problema: lo ridefinisce con maggiore precisione, mostrando che esso riguarda la natura delle relazioni causali costitutive del contenuto, non la distribuzione del contenuto tra interno ed esterno.

Rimane aperta anche la questione dell'applicazione empirica del formalismo: come si misura la curvatura semantica? Come si verifica empiricamente che due stati appartengano alla stessa classe di equivalenza diffeomorfica? Queste domande esigono un programma di ricerca che attraversa la semantica formale, la psicologia cognitiva sperimentale e le neuroscienze computazionali — un programma che questa monografia non ha l'ambizione di completare ma che spera di rendere, almeno, più chiaramente formulabile.

Ciò che il relazionalismo semantico offre, in ultima istanza, non è una soluzione definitiva al problema stretto/largo ma una *diagnostica* della sua struttura e una *riformulazione* che dissolve la dicotomia sostituendola con una geometria. La mente non è dentro, il mondo non è fuori: vi è uno spazio semantico — curvo, dinamico, storicamente stratificato — di cui la mente e il mondo sono le coordinate. Il contenuto è la struttura differenziale di questo spazio. Esplorarne la geometria è, forse, il compito più urgente di una filosofia della mente che voglia restare fedele tanto al rigore formale quanto alla complessità del suo oggetto.

Realismo Semantico e il Problema della Composizionalità Relazionale

---

XXIII. La curvatura semantica come argomento per il realismo sui tipi naturali

Una delle implicazioni più feconde del relazionalismo semantico — e, a mio avviso, tra le meno ovvie — riguarda la metafisica dei tipi naturali. Se il contenuto è una struttura differenziale sullo spazio $S \times E$, e se la curvatura di questa struttura è determinata dalla relazione causale tra il sistema cognitivo e l'ambiente, allora la *forma* della curvatura semantica riflette qualcosa di reale nella struttura dell'ambiente stesso. Non si tratta di un argomento nuovo per il realismo scientifico — la convergenza delle teorie verso gli stessi enti è l'argomento standard — ma di un argomento di tipo diverso, che procede non dall'accordo intersoggettivo ma dalla *geometria interna* dello spazio semantico.

L'argomento si articola come segue. Consideriamo la funzione di contenuto $\mathscr{C}$ per un sistema cognitivo maturo inserito in un ambiente fisicamente strutturato. Le regioni dello spazio E degli ambienti in cui $\mathscr{C}$ mostra una curvatura particolarmente alta — in cui piccole variazioni ambientali producono grandi variazioni di contenuto — corrispondono a *frontiere semanticamente salienti*: confini nell'ambiente rispetto ai quali il contenuto è particolarmente sensibile. La domanda è: perché certe regioni di E sono frontiere salienti piuttosto che altre?

Una risposta puramente convenzionalista direbbe: perché i parlanti hanno scelto, per ragioni pragmatiche o storiche, di tracciare i confini concettuali in certi punti piuttosto che in altri. Ma questa risposta è insufficiente per i tipi naturali. Se il confine semantico tra acqua e non-acqua corrisponde al confine tra H_2O e le sostanze che non lo sono, e se questo confine è semanticamente saliente per sistemi cognitivi di specie diverse, in culture diverse, con storie linguistiche indipendenti, allora la saggezza convenzionalista diventa implausibile: non vi è ragione per cui convenzioni indipendenti convergano sullo stesso confine se quel confine non riflette una struttura reale nell'ambiente.

Formalmente: la saggezza convenzionalista afferma che la distribuzione delle curvature in $\mathscr{C}$ è determinata dalla scelta delle sezioni canoniche σ — dalla scelta degli ambienti normativi rispetto ai quali i contenuti sono valutati. Sistemi cognitivi diversi con sezioni diverse avranno distribuzioni di curvatura diverse, e non vi è convergenza attesa. Il realismo sui tipi naturali afferma invece che vi è una classe privilegiata di sezioni — quelle che si allineano con la struttura causale del mondo — e che i sistemi cognitivi che sopravvivono e si riproducono sono sistemi la cui struttura interna li spinge verso sezioni di questo tipo. La pressione selettiva non è convenzionale ma causale: i sistemi che tracciano confini semantici in corrispondenza di frontiere causali

reali ottengono migliori performance cognitive perché le frontiere causali reali sono quelle che strutturano le regolarità del mondo.

Questo argomento connette il realismo sui tipi naturali alla teoria evolutiva del contenuto in un modo che evita il problema dell'indeterminatezza del livello selettivo discusso nella Parte Quarta. L'obiezione era che la pressione selettiva non individua univocamente il livello di descrizione corretto (H_2O contro il liquido incolore che disseta). Ma se si argomenta che la struttura della curvatura semantica converge sui confini causali profondi per ragioni strutturali — perché sono quelli che strutturano le regolarità predittivamente più potenti — allora la teleosemantica non deve individuare il livello giusto per convenzione: il livello giusto è quello in cui la curvatura semantica è massimalmente allineata con la struttura causale del mondo.

La distinzione tra "il livello di H_2O" e "il livello del liquido incolore che disseta" non è arbitraria se si considera la struttura delle regolarità causali: le regolarità che coinvolgono H_2O (reazioni chimiche, proprietà termiche, comportamento in presenza di diversi soluti) sono molto più profonde e sistematiche delle regolarità che coinvolgono "il liquido incolore che disseta". Un sistema cognitivo che tracci il confine a livello di H_2O ha accesso a una rete di regolarità predittive più densa e coerente di uno che lo tracci al livello fenomenologico. La curvatura semantica al livello di H_2O è massimalmente allineata con la struttura causale del mondo: è per questo che essa emerge come la sezione canonica privilegiata per i sistemi cognitivi immersi in quell'ambiente.

Questo argomento non è un argomento per il realismo ingenuo: non afferma che i tipi naturali esistano indipendentemente da qualsiasi sistema cognitivo, in una forma platonica pura. Afferma qualcosa di più modesto e più difendibile: che vi sono strutture causali nel mondo — frontiere tra regimi di regolarità — e che i sistemi cognitivi efficaci sono quelli la cui funzione di contenuto $\mathscr{C}$ è sensibile a queste frontiere. I tipi naturali non sono enti misteriosi: sono le strutture dell'ambiente lungo le cui frontiere la curvatura semantica di sistemi cognitivi evolutivamente adeguati è massima.

XXIV. Il problema della composizionalità nel quadro relazionale

La composizionalità del contenuto — il fatto che il contenuto di un'espressione complessa sia sistematicamente determinato dal contenuto delle sue parti e dalla struttura sintattica con cui si combinano — è uno dei vincoli più stringenti che qualsiasi teoria semantica deve soddisfare. Il principio di composizionalità è ciò che spiega come esseri cognitivi finiti possano comprendere un'infinità di espressioni nuove: non occorre imparare il contenuto di ogni espressione caso per caso, perché il contenuto dei complessi si calcola compositivamente a partire dai componenti noti.

Il relazionalismo semantico incontra qui una difficoltà non banale. Se il contenuto di uno stato mentale è la struttura differenziale della funzione $\mathscr{C}$ in un punto $(s, e) \in S \times E$, com'è definita la "composizione" di due stati mentali nel quadro relazionalista? Come si combinano le strutture differenziali locali di due funzioni di contenuto per produrre la struttura differenziale del loro composto?

Il problema emerge in tutta la sua acuità quando si consideri che la composizionalità standard opera sui *contenuti* — sulle proposizioni o sui contributi proposizionali — trattandoli come entità composibili secondo operazioni algebriche. Il contenuto della proposizione "l'acqua è bagnata" è determinato dal contenuto di "acqua" e dal contenuto di "è bagnata" combinati secondo le regole della semantica vero-condizionale. Ma nel quadro relazionalista, il contenuto non è un oggetto fisso: è una struttura locale che dipende dal punto in $S \times E$. Come si compongono strutture locali?

Propongo di affrontare questo problema introducendo la nozione di *composizione funtoriale* del contenuto. L'idea è che la composizionalità non operi direttamente sui valori di $\mathscr{C}$ ma sulla struttura delle funzioni di contenuto — che vi sia un'operazione di composizione a livello delle funzioni, non dei valori. Formalmente, se $\mathscr{C}_\alpha$ e $\mathscr{C}_\beta$ sono le funzioni di contenuto per due stati mentali componenti (corrispondenti, diciamo, ai concetti di "acqua" e di "è bagnata"), la funzione di contenuto del pensiero composto non è semplicemente $\mathscr{C}_\alpha \otimes \mathscr{C}_\beta$ per qualche operazione $\otimes$ sulle proposizioni: è una funzione $\mathscr{C}_{[\alpha,\beta]} : S_comp \times E \to P_comp$ che dipende da uno spazio degli stati composito S_comp e da uno spazio delle proposizioni composito P_comp.

La struttura di S_comp non è il mero prodotto $S_\alpha \times S_\beta$: è uno spazio che incorpora le *relazioni* tra gli stati componenti, le connessioni sintattiche e semantiche che li legano nella struttura composita. Formalmente, S_comp è un *fibrato* su $S_\alpha \times S_\beta$, dove la fibra in ogni punto (s_α, s_β) cattura le modalità di combinazione sintattica:

$$S_comp = \{ (s_\alpha, s_\beta, \varrho) : s_\alpha \in S_\alpha, s_\beta \in S_\beta, \varrho \in R(s_\alpha, s_\beta) \}$$

dove $R(s_\alpha, s_\beta)$ è lo spazio delle relazioni sintattico-semantiche ammissibili tra gli stati s_α e s_β. Questa struttura di fibrato permette di catturare il fatto che la composizione non è libera: non tutti i piani di stati componenti possono essere combinati in qualsiasi modo, ma solo secondo modalità strutturalmente coerenti con le rispettive funzioni di contenuto.

La funzione di contenuto composita è allora:

$$\mathscr{C}_{[\alpha,\beta]}(s_\alpha, s_\beta, \varrho, e) = \Phi_\varrho(\mathscr{C}_\alpha(s_\alpha, e), \mathscr{C}_\beta(s_\beta, e))$$

dove Φ_ϱ è l'operazione di composizione semantica determinata dalla relazione sintattica ϱ — per esempio, l'applicazione funzionale nella semantica di tipo teorico, o la predicazione nella semantica relazionale classica. Il fatto cruciale è che Φ_ϱ dipende

da ϱ — dalla struttura sintattica che connette i componenti — e che ϱ è a sua volta un elemento dello spazio degli stati compositi, non un parametro esterno alla funzione di contenuto.

Questo schema ha due proprietà di notevole interesse. La prima è che la composizionalità relazionale è *locale*: la funzione di contenuto del composto è determinata dalle funzioni di contenuto dei componenti e dalla struttura sintattica, ma questa determinazione è locale nel senso che dipende dalla struttura differenziale locale delle funzioni componenti, non dai loro valori in punti lontani. La seconda proprietà è che la composizionalità relazionale è *sensibile al contesto*: poiché le funzioni di contenuto componenti dipendono da e, e la composizione avviene preservando questa dipendenza, il contenuto del composto eredita la sensibilità ambientale dei suoi componenti. La composizionalità non elimina l'esternalismo: lo incorpora.

XXV. La composizionalità relazionale e il problema del contesto

Questa struttura funtoriale permette di affrontare un problema che affligge qualsiasi teoria composizionale: la *dipendenza contestuale* del contenuto dei costituenti. È ben noto che molti costituenti espressivi — gli aggettivi in particolare — non hanno un contenuto fisso ma uno che varia sistematicamente con il contesto in cui appaiono. "Alto" significa cose diverse se applicato a un bambino, a un adulto o a un palazzo. "Veloce" varia il suo contenuto con il tipo di oggetto cui si applica. La semantica di tipo teorico standard gestisce questo fenomeno con funzioni di tipo $\langle e,t\rangle \to \langle e,t\rangle$ che modificano predicati, ma questo approccio è puramente estensionale e non cattura la struttura della dipendenza contestuale.

Nel quadro relazionalista, la dipendenza contestuale degli aggettivi corrisponde a una *curvatura incrociata* nella funzione di contenuto composita: la derivata parziale della funzione di contenuto di "alto bambino" rispetto alla variabile sintattica ϱ (che codifica il tipo dell'oggetto a cui l'aggettivo si applica) non è nulla. Il contenuto di "alto" nel composto "bambino alto" non è il valore di $\mathscr{C}$_alto(s_alto, e) valutato indipendentemente e poi combinato col contenuto di "bambino": è il valore di $\mathscr{C}$_alto nella fibra R(s_alto, s_bambino), dove la struttura della fibra incorpora l'informazione sul tipo dell'oggetto.

Formalmente:

$\partial \mathscr{C}$_[alto, bambino]$/\partial \varrho \neq 0$

La non-nullità di questa derivata cattura il fatto che la struttura sintattica — il tipo dell'oggetto cui l'aggettivo è applicato — co-determina il contenuto dell'intero composto, non semplicemente per applicazione di una funzione dal contenuto fisso

dell'aggettivo, ma per deformazione della funzione di contenuto dell'aggettivo lungo la fibra sintattica.

Questo suggerisce che la composizionalità relazionale non è composizionalità nel senso standard — determinazione del contenuto del composto dai contenuti fissi dei componenti e dalla struttura sintattica — ma qualcosa di più sottile: *determinazione della struttura differenziale del composto dalle strutture differenziali dei componenti lungo le fibre sintattiche*. La composizionalità è preservata, ma come proprietà delle strutture differenziali, non dei valori.

La differenza è filosoficamente rilevante perché il senso standard di composizionalità tende a postulare contenuti atomici fissi — i contenuti dei componenti minimi — da cui i contenuti complessi si costruiscono. Questa postulazione ha il sapore di un assunto internalista: i contenuti atomici sono trattati come proprietà fisse degli stati mentali, indipendenti dal contesto composizionale. La composizionalità relazionale abbandona questo assunto: i contenuti atomici non sono fissi ma variano con il contesto composizionale, e questa variazione è strutturata dalla geometria delle fibre sintattiche. Il contenuto di "alto" non esiste come entità atomica indipendente: esiste come struttura differenziale locale che prende valore solo in un contesto composizionale specificato.

Questa conclusione ha un'eco nella semantica contestuale di Recanati e nella pragmatica dell'arricchimento proposizionale: il contenuto esplicito delle espressioni non è determinato dalla sola semantica composizionale ma richiede un'integrazione pragmatica che specifichi il contesto. Il relazionalismo semantico offre una formalizzazione di questo processo di integrazione: il contesto non è un parametro esterno alla semantica ma è incorporato nella struttura delle fibre sintattiche e nella dipendenza di e dello spazio degli ambienti.

XXVI. Il problema dell'opacità referenziale nel quadro relazionale

Un banco di prova classico per qualsiasi teoria del contenuto è il trattamento dei contesti intenzionali opachi — i contesti in cui la sostituzione di termini coreferenciali non preserva il valore di verità. "Edipo vuole sposare Giocasta" è vera; "Edipo vuole sposare sua madre" ha lo stesso contenuto putativo, poiché Giocasta è sua madre; eppure, nella lettura intenzionale, la seconda frase non è equivalente alla prima — Edipo non ha la credenza di voler sposare sua madre. L'opacità referenziale segnala che i contesti di credenza e desiderio non sono sensibili soltanto al riferimento dei termini che vi compaiono, ma al *modo* in cui il riferimento è dato al soggetto.

La struttura fregiana del Sinn — del senso come modo di presentazione dell'oggetto — è la risposta classica a questo fenomeno: i contesti intenzionali sono sensibili al senso, non al riferimento. Ma nel quadro relazionalista, questa risposta richiede una

riformulazione, perché il "senso" non è un'entità mentale interna (contro Frege psicologisticamente letto) né un'entità astratta platonica (Frege ortodosso), ma una struttura differenziale locale della funzione di contenuto.

Propongo che l'opacità referenziale corrisponda alla *non-invarianza della funzione di contenuto sotto sostituzione di termini coreferenciali lungo le direzioni sintattiche*. Due termini τ_1 e τ_2 sono coreferenciali se $\mathscr{C}(s_\tau_1, e) = \mathscr{C}(s_\tau_2, e)$ per tutti gli ambienti e rilevanti — se producono lo stesso valore nella funzione di contenuto. Ma la sostituzione di τ_1 con τ_2 in un contesto composito non preserva il contenuto del composto se le *strutture differenziali locali* di $\mathscr{C}_\tau_1$ e $\mathscr{C}_\tau_2$ sono diverse: se le due funzioni di contenuto hanno curvature locali diverse, anche producendo gli stessi valori, la composizione con altri elementi produrrà strutture composite diverse.

Formalmente: la sostituzione $\tau_1 \rightarrow \tau_2$ in un composto $[\alpha, \tau]$ preserva il contenuto del composto se e solo se:

$$\mathscr{C}_[\alpha, \tau_1] = \mathscr{C}_[\alpha, \tau_2]$$

Questa condizione non è equivalente alla coreferencialità di τ_1 e τ_2 (uguale valore di $\mathscr{C}$ su E), ma richiede l'uguaglianza delle strutture differenziali locali:

$$d\mathscr{C}_\tau_1 = d\mathscr{C}_\tau_2 \text{ su tutto l'intorno rilevante}$$

"Giocasta" e "la madre di Edipo" possono essere coreferenciali — produrre lo stesso valore di $\mathscr{C}$ per gli ambienti reali — ma avere differenziali locali diversi: il termine "Giocasta" ha una funzione di contenuto la cui dipendenza da e è strutturata dalla catena causale che lega Edipo alla donna di nome Giocasta; il termine "la madre di Edipo" ha una funzione di contenuto la cui dipendenza da e è strutturata dalla relazione di maternità. Queste due strutture causali sono diverse — producono curvature diverse in regioni controfattuali dello spazio E — e perciò, quando si compongono con la funzione di contenuto di "Edipo vuole sposare", producono strutture composite diverse.

L'opacità referenziale è dunque, nel quadro relazionalista, il fenomeno della *sensibilità dei composti alla struttura differenziale locale dei componenti*, non soltanto ai loro valori. Questo risolve il problema senza dover postulare i sensi come entità aggiuntive: la struttura differenziale locale della funzione di contenuto è il "senso" nel quadro relazionalista — non un oggetto aggiunto alla teoria, ma una proprietà geometrica delle funzioni di contenuto che il formalismo già possedeva.

La connessione con la distinzione de re/de dicto diventa particolarmente nitida in questa luce. Una credenza de re è una credenza il cui contenuto dipende direttamente dall'individuo referito — è sensibile alla componente posizionale di E più che alla struttura interna del termine. Una credenza de dicto è una credenza il cui contenuto dipende dal modo di presentazione — è sensibile alla struttura differenziale locale

della funzione di contenuto, non soltanto al suo valore referenziale. Nel formalismo relazionalista:

Credenza de re: $\partial\mathscr{C}/\partial e_P$ è la derivata dominante
Credenza de dicto: $d\mathscr{C}$ locale (struttura differenziale) è la proprietà dominante

Le credenze miste — de re rispetto a un argomento, de dicto rispetto a un altro — corrispondono a composti in cui diversi costituenti hanno profili di derivate parziali diversi: alcuni dominati dalla componente posizionale, altri dalla struttura differenziale locale. Questa pluralità di profili in un unico composto spiega la complessità dei contesti intensionali misti, dove l'interpretazione oscilla tra letture de re e de dicto per diversi costituenti dello stesso pensiero.

Autoconoscenza, Trasparenza e il Paradosso del Contenuto Opaco a Se Stessi

---

XXVII. Il privilegio in prima persona e la sua fragilità relazionalista

Uno degli assunti più radicati nella tradizione filosofica è che il soggetto abbia un accesso privilegiato ai propri stati mentali — che la conoscenza di ciò che si pensa o si crede non richieda inferenza, osservazione o indagine, ma sia data in modo immediato e autoreferenziale. Questo privilegio in prima persona ha ricevuto formulazioni diverse nel corso della storia: dalla cogitatio cartesiana all'appercezione kantiana, dalla certezza husserliana dei vissuti alla trasparenza del contenuto mentale propugnata da Evans e McDowell nella filosofia analitica contemporanea. Al di là delle differenze di vocabolario, vi è un nucleo comune: il soggetto sa, o è in posizione privilegiata per sapere, *che cosa* pensa.

Il relazionalismo semantico mina questo privilegio in modo non banale, e lo fa attraverso una conseguenza che emerge direttamente dalla struttura della funzione di contenuto $\mathscr{C}$. Se il contenuto di uno stato mentale è la struttura differenziale di $\mathscr{C}$ in un punto $(s, e) \in S \times E$, e se la componente e di questo punto è determinata dall'ambiente esterno — da fatti causali che trascendono la prospettiva in prima persona del soggetto —, allora il soggetto non ha, in linea di principio, accesso pieno al contenuto dei propri stati mentali. Egli ha accesso diretto alla componente s — allo stato interno — ma non alla componente e, e quindi non alla struttura differenziale completa che costituisce il contenuto.

Questa conclusione non è nuova nella sua formulazione generica: Burge l'ha difesa nella sua analisi dell'autoconoscenza, mostrando che i soggetti possono sbagliarsi sul contenuto delle proprie credenze quando quel contenuto è determinato da norme

sociali che il soggetto non padroneggia completamente. Ma nel quadro relazionalista la difficoltà è più profonda e sistematica di quanto la versione burghiana non suggerisca. Non si tratta solo che il soggetto ignori le norme sociali: si tratta che la struttura differenziale completa di $\mathscr{C}$ è intrinsecamente inaccessibile dalla prospettiva in prima persona, non per mancanza di informazione contingente, ma per ragioni strutturali.

La ragione strutturale è la seguente. Il soggetto, dalla propria prospettiva interna, ha accesso a ciò che potremmo chiamare la *proiezione interna* della funzione di contenuto: la funzione parziale $\mathscr{C}(\cdot, e_attuale)$ ottenuta fissando l'ambiente all'ambiente reale in cui il soggetto si trova. Questa proiezione è la funzione che il soggetto percepisce come "il suo contenuto" — ma è appunto una proiezione, non la funzione di contenuto completa. La struttura differenziale completa di $\mathscr{C}$ — la forma del differenziale totale $d\mathscr{C}$ con entrambe le derivate parziali — richiede la conoscenza di come il contenuto vari non solo al variare degli stati interni ma anche al variare dell'ambiente. E questa conoscenza non è disponibile dalla prospettiva in prima persona, perché richiede la conoscenza controfattuale di ciò che si penserebbe in ambienti diversi — conoscenza che trascende il punto di vista interno.

Il soggetto, insomma, vive nella fibra: ha accesso alla struttura verticale della funzione di contenuto — come essa varia con lo stato interno a parità di ambiente — ma non alla struttura orizzontale — come essa varia con l'ambiente a parità di stato interno. Il contenuto pieno è la struttura dell'intera fibrazione, e la fibra su cui il soggetto vive è solo una sezione di essa.

XXVIII. La trasparenza di Evans e la sua riformulazione relazionale

Gareth Evans, nel postumo *The Varieties of Reference* (1982), ha difeso una tesi di *trasparenza* del contenuto mentale che merita un esame ravvicinato, perché la sua plausibilità superficiale occulta presupposti che il relazionalismo semantico può smontare con precisione. La tesi di Evans è che i pensieri dotati di contenuto singolare — i pensieri su oggetti individuali, le credenze de re — richiedono che il soggetto sia in una relazione informazionale appropriata con l'oggetto del pensiero, e che questa relazione sia tale da rendere l'oggetto immediatamente disponibile alla coscienza del soggetto. Un pensiero su x richiede che x sia, in un senso preciso, "dato" al soggetto, non mediato da descrizioni che potrebbero non cogliere l'individuo corretto.

La tesi evansoniana ha un'apparente affinità con l'esternalismo: riconosce che il contenuto singolare dipende dalla relazione del soggetto con l'oggetto, non solo dagli stati interni. Ma Evans è anche sostenitore del privilegio in prima persona: il soggetto che ha un pensiero su x sa di avere un pensiero su x — non perché ne faccia oggetto di esame esterno, ma perché la relazione informazionale con x è constitutiva del pensiero stesso e perciò trasparente a chi lo ha.

Nel quadro relazionalista, la tesi di Evans corrisponde all'ipotesi che la proiezione interna di $\mathscr{C}$ sulla fibra e_attuale sia *fedele* rispetto all'intera struttura di $\mathscr{C}$ nelle sue proprietà semanticamente cruciali. Ossia: che la conoscenza di $\mathscr{C}(\cdot, \text{e_attuale})$ — la conoscenza "dall'interno" del contenuto — preservi le proprietà che contano per l'identità del pensiero, anche senza conoscenza della struttura orizzontale completa.

Questa ipotesi di fedeltà non è globalmente falsa, ma è localmente limitata: vale per certi aspetti del contenuto (il riferimento dell'oggetto, la predica applicata) ma non per altri (la struttura controfattuale del contenuto, la sua sensibilità a variazioni ambientali non attuali). Evans ha ragione che il soggetto sa *che cosa* pensa nel senso del valore referenziale del pensiero — chi è l'oggetto del pensiero nell'ambiente attuale. Ma non ha accesso alla curvatura semantica del pensiero — a come esso varierebbe in ambienti controfattuali, a quali frontiere ambientali lo renderebbero inappropriato o lo farebbero cambiare contenuto. Il privilegio in prima persona è reale ma *parziale*: copre la proiezione attuale del contenuto, non la sua struttura differenziale piena.

Questa parzialità ha conseguenze precise. Consideriamo il caso di un soggetto che ha sviluppato, nel corso di una vita, una rappresentazione interna dell'acqua — una funzione di contenuto $\mathscr{C}$_acqua — calibrata sull'ambiente terrestre di H_2O. Se questo soggetto fosse trasferito sulla Terra Gemella, la proiezione interna della sua funzione di contenuto sull'ambiente di XYZ gli darebbe un pensiero che si sente dall'interno identico al pensiero "acqua" — ma che ha ora un contenuto parzialmente diverso rispetto a quello che aveva sulla Terra. Il soggetto non saprebbe, dalla prospettiva interna, che il contenuto del suo pensiero si è modificato. La trasparenza evansoniana fallisce proprio per i casi controfattuali che rivelano la struttura differenziale del contenuto: il soggetto è trasparente rispetto alla proiezione attuale, opaco rispetto alla struttura globale.

XXIX. L'autoconoscenza come ricostruzione parziale della curvatura

Questa situazione non implica che l'autoconoscenza sia impossibile o globalmente inaffidabile: implica che essa abbia una struttura più complessa di quella che la tradizione della trasparenza postula. L'autoconoscenza non è accesso diretto alla struttura differenziale piena di $\mathscr{C}$: è un processo di *ricostruzione parziale* di quella struttura a partire da dati accessibili in prima persona.

Quali dati sono disponibili? Il soggetto ha accesso diretto alla proiezione interna $\mathscr{C}(\cdot, \text{e_attuale})$ — sa, per gli ambienti che conosce, quali contenuti i suoi stati mentali producono. Ha accesso parziale alle sue disposizioni inferenziali — alle transizioni tra stati mentali che il suo sistema cognitivo compie — e queste disposizioni riflettono, indirettamente, la struttura di $\mathscr{C}$ nelle sue direzioni interne (le derivate $\partial\mathscr{C}/\partial s$). Ha accesso ai propri errori semantici passati, quando li vengono notificati dalla comunità

linguistica o dall'esperienza — e questi errori rivelano, caso per caso, aspetti della struttura orizzontale di $\mathscr{C}$ che non erano accessibili dall'interno.

L'autoconoscenza del contenuto è dunque una forma di *inferenza al miglior agente* della struttura differenziale di $\mathscr{C}$ a partire da dati necessariamente parziali. Non è accesso diretto ma costruzione teorica — nel senso in cui un fisico ricostruisce la geometria di uno spazio curvo a partire da misurazioni locali, senza mai avere accesso diretto alla struttura globale. Il soggetto cognitivo è, rispetto ai propri contenuti, nella posizione di un geometra che abita la superficie che cerca di misurare: può determinare la curvatura locale con precisione arbitrariamente alta, ma la struttura globale rimane accessibile solo attraverso una rete di misurazioni locali e la loro coerenza teorica.

Questa concezione dell'autoconoscenza ha implicazioni per il dibattito tra *internalismo epistemico* e *esternalismo epistemico* sull'autoconoscenza. L'internalismo epistemico sostiene che le credenze di un soggetto sui propri stati mentali sono giustificate da fattori accessibili al soggetto stesso — dalla sua prospettiva in prima persona. L'esternalismo epistemico sostiene che la giustificazione delle credenze autoreferenziali può dipendere da fattori esterni alla prospettiva soggettiva. Il relazionalismo semantico suggerisce che entrambe le posizioni catturano aspetti reali: l'internalismo epistemico è adeguato rispetto alla proiezione interna del contenuto (l'accesso alla fibra verticale), l'esternalismo epistemico è necessario per la struttura orizzontale (la curvatura rispetto ad e). La giustificazione dell'autoconoscenza è *stratificata* come il contenuto stesso.

XXX. Il paradosso del contenuto opaco a se stessi: una diagnosi

Il paradosso che il relazionalismo semantico produce sull'autoconoscenza può essere formulato in modo acuto come segue. Da un lato, i contenuti mentali sono ciò che il soggetto *pensa* — sono l'oggetto diretto e immediato dell'attività cognitiva, non qualcosa di cui il soggetto fa esperienza come di una realtà esterna. Dall'altro, il relazionalismo mostra che la struttura completa di questi contenuti — la loro curvatura semantica, la loro sensibilità controfattuale, la loro collocazione nello spazio $S \times E$ — è parzialmente opaca al soggetto dalla propria prospettiva interna. Come può essere che ciò che penso mi sia in parte opaco?

Il paradosso si dissolve non negando uno dei due corni ma distinguendo due sensi di "opacità". Vi è un'opacità di *accesso* — l'incapacità di avere accesso diretto a certi aspetti del proprio contenuto — e vi è un'opacità di *costituzione* — il fatto che il proprio contenuto sia costituzionalmente dipendente da fattori che trascendono la prospettiva interna. Queste due forme di opacità sono logicamente indipendenti: si può avere accesso indiretto e ricostruttivo a ciò che è costituzionalmente esterno; si può non avere accesso diretto a qualcosa che è costituzionalmente interno (i processi inconsci, per esempio).

Nel caso del contenuto relazionalista, si tratta di un'opacità di accesso derivante da opacità di costituzione: il contenuto è costituzionalmente determinato da fattori ambientali cui il soggetto non ha accesso diretto, e quindi la struttura completa del contenuto non è accessibile dall'interno. Ma questo non è un paradosso: è una caratteristica della condizione epistemica di qualsiasi agente incarnato in un mondo che non controlla. Il soggetto cognitivo è in un mondo che co-costituisce i suoi pensieri; non può stare fuori da questo processo di co-costituzione per osservarlo neutralmente; può soltanto muoversi all'interno di esso con l'ausilio delle risorse ricostruttive — l'inferenza, la comunicazione, la correzione intersoggettiva — che la sua condizione gli mette a disposizione.

Il privilegio in prima persona non è soppresso dal relazionalismo semantico: è *ridimensionato e ricollocato*. Il soggetto ha un accesso privilegiato alla proiezione del proprio contenuto sull'ambiente attuale — sa meglio di qualsiasi osservatore esterno come il mondo gli appare, quale struttura ha la propria esperienza presente. Ma non ha accesso privilegiato alla struttura differenziale completa del proprio contenuto — alla forma dello spazio semantico che abita. Questa asimmetria tra accesso alla fibra e accesso alla struttura globale è la versione relazionalista della distinzione classica tra il *come mi appare* (accessibile in prima persona) e il *che cosa penso davvero* (parzialmente trascendente la prima persona). Non è una versione scettica di questa distinzione — non si afferma che il soggetto sia sistematicamente ingannato rispetto ai propri pensieri — ma una versione strutturalmente informata: il grado di trasparenza del contenuto è una proprietà locale della geometria dello spazio semantico, determinabile in linea di principio per ciascun tipo di contenuto, non una proprietà globale che si afferma o si nega en bloc.

Il Relazionalismo nello Spazio Modale: Mondi Possibili, Contenuto a Due Dimensioni e la Geometria dell'Intensionalità

XXXI. La semantica dei mondi possibili come sfida e come risorsa

La semantica dei mondi possibili costituisce lo strumento tecnico più potente di cui la filosofia analitica dispone per trattare il contenuto proposizionale, la necessità, la possibilità e la dipendenza controfattuale. Nella sua formulazione standard, un contenuto proposizionale è una funzione caratteristica da un insieme di mondi possibili W all'insieme dei valori di verità {V, F}: una proposizione è vera in certi mondi e falsa in altri, e l'insieme dei mondi in cui è vera — la sua *estensione intensionale* — individua il suo contenuto. Due proposizioni con la stessa estensione intensionale — vere negli stessi mondi — hanno lo stesso contenuto; due proposizioni con estensioni diverse hanno contenuti diversi.

Questo framework è potente ma incontra difficoltà ben note per i contenuti intenzionali. Il problema delle *proposizioni necessariamente equivalenti* è il più imbarazzante: tutte le tautologie logiche sono vere in tutti i mondi possibili e hanno perciò la stessa estensione intensionale, ma è chiaramente sbagliato dire che pensare "p ∨ ¬p" e pensare "q ∨ ¬q" (per p ≠ q) sia pensare la stessa cosa. Il problema delle *proposizioni matematicamente necessarie* è analogo: "7 + 5 = 12" e "il quadrato dell'ipotenusa è uguale alla somma dei quadrati dei cateti" sono entrambe necessariamente vere e hanno la stessa estensione intensionale, ma il pensiero che le esprime è cognitivamente molto diverso.

La semantica bidimensionale — sviluppata da Davies e Humberstone, poi elaborata da Chalmers con grande sofisticazione — risponde a queste difficoltà introducendo due dimensioni di valutazione invece di una. Nella semantica bidimensionale standard, un'espressione è valutata rispetto a una coppia di mondi possibili ⟨w_c, w_e⟩, dove w_c è il *mondo considerato come attuale* (il contesto, che determina il riferimento dei termini indicali e dei designatori rigidi) e w_e è il *mondo considerato come controfattuale* (la circostanza di valutazione, rispetto alla quale si determina il valore di verità). Il contenuto di un'espressione ha perciò due componenti: l'*intenzione primaria* (che varia con w_c, catturando il significato epistemico o a priori) e l'*intenzione secondaria* (che varia con w_e a contesto fisso, catturando il significato metafisico o a posteriori).

Il relazionalismo semantico ha un rapporto complesso con la semantica bidimensionale: ne condivide la motivazione — riconoscere che il contenuto ha una struttura più ricca di quanto la semplice estensione intensionale catturi — ma si discosta da essa su punti strutturalmente importanti. Voglio sviluppare questa divergenza con precisione, perché essa illumina tanto la potenza quanto i limiti del relazionalismo rispetto agli strumenti formali disponibili.

XXXII. Intenzioni primarie, secondarie e struttura differenziale: un confronto

Nella semantica bidimensionale di Chalmers, l'intenzione primaria di un termine come "acqua" è approssimativamente: *il liquido incolore, insapore, che scorre nei fiumi, disseta, ecc., di qualunque tipo esso sia nell'ambiente in cui questo contesto si svolge*. L'intenzione primaria è una funzione da W a estensioni: per ogni mondo w considerato come attuale, restituisce l'insieme di oggetti che soddisfano le condizioni fenomenologiche e funzionali associate al termine in quel mondo. L'intenzione secondaria, fissato il mondo attuale come la Terra (dove l'acqua è H_2O), è: *l'insieme di tutte le cose che sono H_2O in qualsiasi mondo controfattuale*. L'intenzione secondaria è rigida rispetto al tipo naturale: individua H_2O in tutti i mondi, anche quelli in cui H_2O non scorre nei fiumi né disset a.

La struttura bidimensionale captura, nella sua architettura, la medesima intuizione che motiva il relazionalismo: il contenuto dipende sia dall'ambiente del soggetto (il contesto, w_c) sia dalla struttura del mondo valutato (la circostanza, w_e). La coppia

⟨w_c, w_e⟩ è l'analogo modale del punto (s, e) ∈ S × E nel relazionalismo. Le intenzioni primaria e secondaria sono, rispettivamente, le proiezioni della funzione di contenuto lungo le due dimensioni del prodotto.

La divergenza emerge tuttavia su un punto decisivo: la semantica bidimensionale tratta W come un insieme fisso di mondi possibili, indipendente dal soggetto che valuta il contenuto. Il relazionalismo semantico, per contro, insiste che la struttura dello spazio E degli ambienti rilevanti non è fissa ma dipende dal sistema cognitivo in esame: la curvatura di $\mathscr{C}$ su S × E — la sensibilità del contenuto a certi tipi di variazione ambientale — è una proprietà del sistema, non uno sfondo neutro. Due sistemi cognitivi diversi, inseriti nello stesso ambiente, possono avere funzioni di contenuto $\mathscr{C}$ con curvature diverse: uno è semanticamente sensibile alla distinzione H_2O/XYZ, l'altro non lo è, perché la sua storia causale e la sua struttura cognitiva non lo hanno reso sensibile a quella frontiera.

Questa differenza ha conseguenze formali precise. Nella semantica bidimensionale, le intenzioni primaria e secondaria di un termine sono definite rispetto all'intero spazio W: ogni soggetto con competenza sul termine "acqua" ha la stessa coppia di intenzioni, perché le intenzioni sono proprietà del termine (o dell'idioletto), non del soggetto come individuo fisico-cognitivo. Nel relazionalismo, la funzione di contenuto $\mathscr{C}$ è indicizzata al sistema cognitivo: diversi sistemi possono avere diverse funzioni $\mathscr{C}$, e la struttura dello spazio S × E è soggetto-relativa nel senso precisamente determinato dalla storia causale e dalla struttura interna del sistema.

Formalmente, la semantica bidimensionale definisce l'intenzione primaria come:

IP_τ : W → Ext(τ)
IP_τ(w) = { x ∈ D(w) : x soddisfa le condizioni primarie di τ in w }

e l'intenzione secondaria come:

IS_τ : W → Ext(τ)
IS_τ(w) = { x ∈ D(w) : x è dello stesso tipo di IP_τ(w_att) in w }

dove w_att è il mondo attuale. Il relazionalismo propone invece:

$\mathscr{C}$_τ^A : S_A × E → P
$\mathscr{C}$_τ^A(s, e) = contenuto di τ per il sistema A con stato interno s in ambiente e

dove il sopraindice A indica la dipendenza dal sistema cognitivo specifico. La distinzione strutturale è che $\mathscr{C}$_τ^A dipende da S_A — dallo spazio degli stati interni del sistema A, che è una proprietà fisica e storica del sistema — mentre le intenzioni kaplaniane o chalmersiane non dipendono dal sistema fisico ma soltanto dal termine o idioletto.

XXXIII. La necessità a posteriori nel quadro relazionale

La necessità a posteriori — la classe di verità necessarie ma conoscibili solo empiricamente, la cui esistenza è stata rivelata da Kripke — costituisce uno dei risultati più destabilizzanti della filosofia del linguaggio del Novecento. Che H_2O sia acqua non è una verità contingente: è necessaria, perché acqua *è* H_2O, e questo vale in tutti i mondi possibili in cui l'acqua esiste. Eppure non è conoscibile a priori: richiedeva scoperta empirica, e il contrario era concepibile prima della scoperta. La necessità a posteriori sembra violare la simmetria classica tra necessità e apriorità: non tutto ciò che è necessario è a priori, non tutto ciò che è a posteriori è contingente.

La semantica bidimensionale gestisce questa situazione distinguendo due nozioni di possibilità: la *possibilità primaria* (relativa all'intenzione primaria) e la *possibilità secondaria* (relativa all'intenzione secondaria). È primariamente possibile che l'acqua non sia H_2O — nel senso che vi è un mondo w in cui IP_acqua(w) individua una sostanza diversa da H_2O — ma è secondariamente impossibile, perché l'intenzione secondaria di "acqua" è rigidamente H_2O in tutti i mondi. La concepibilità della non-identità acqua/H_2O corrisponde alla possibilità primaria; la necessità dell'identità corrisponde all'impossibilità secondaria.

Nel quadro relazionalista, la necessità a posteriori riceve una caratterizzazione geometrica naturale. La verità "H_2O è acqua" esprime una proprietà della funzione di contenuto $\mathscr{C}$: il fatto che, nell'ambiente attuale e_att, i termini "H_2O" e "acqua" abbiano la stessa estensione — producano gli stessi valori di verità per ogni stato di cose. Questa co-estensione è *necessaria* nel senso che vale in tutti gli ambienti fisicamente possibili che condividono la struttura chimica dell'ambiente attuale — in tutti i mondi con le stesse leggi fisiche e la stessa chimica. È *a posteriori* nel senso che non è determinata dalla struttura differenziale locale di $\mathscr{C}$ in un intorno arbitrariamente piccolo del punto (s, e_att): richiede la conoscenza della struttura chimica dell'ambiente, che si estende su scala molecolare non accessibile percettivamente.

Formalmente: la necessità a posteriori corrisponde a una *coincidenza globale* dei valori di due funzioni di contenuto su una regione ampia di E, non deducibile dalla struttura locale di $\mathscr{C}$ in nessun suo punto. Si tratta di un'informazione sulla topologia globale dello spazio semantico — sulla struttura di grandi regioni di E — che non è deducibile dalla curvatura locale. Questo spiega perché la necessità a posteriori sia a posteriori: la curvatura locale di $\mathscr{C}$ non la rivela; occorre un'esplorazione sistematica di ampie regioni dello spazio semantico, che è precisamente ciò che la chimica empirica compie.

La distinzione tra necessità e apriorità, nel quadro relazionalista, diventa dunque: la necessità è una proprietà *globale* (topologica) dello spazio semantico — una coincidenza di valori su ampie regioni di E; l'apriorità è una proprietà *locale*

(differenziale) — deducibile dalla struttura di $\mathscr{C}$ in un intorno del punto. Le verità a priori necessarie (come le verità logiche e matematiche) sono quelle la cui necessità è anche localmente determinata: la struttura differenziale di $\mathscr{C}$ le rivela senza richiedere esplorazione globale. Le verità a posteriori necessarie sono quelle la cui necessità globale eccede la struttura locale: rivelabili solo attraverso l'esplorazione empirica dello spazio ambientale.

XXXIV. Il problema di Frege nello spazio modale relazionale

Il problema fregeano delle proposizioni cognitivamente diverse ma necessariamente equivalenti — il problema che la semantica dei mondi possibili standard non può risolvere senza strumenti aggiuntivi — trova nel relazionalismo una soluzione che va nella stessa direzione della semantica bidimensionale ma con una struttura diversa.

Due proposizioni logicamente equivalenti — come "Espero è Espero" e "Espero è Fosforo" (dove Espero e Fosforo sono lo stesso pianeta, Venere) — hanno la stessa estensione intensionale standard: entrambe vere in tutti i mondi in cui Venere esiste. Ma il pensiero che le esprime è cognitivamente diverso: la seconda è una scoperta astronomica, la prima una tautologia. La semantica bidimensionale risolve questo distinguendo le intenzioni primarie delle due proposizioni: l'intenzione primaria di "Espero è Espero" è vera in tutti i mondi (è a priori), mentre quella di "Espero è Fosforo" è vera solo nei mondi in cui il pianeta della sera e il pianeta del mattino sono lo stesso (è a posteriori).

Nel quadro relazionalista, la differenza cognitiva tra i due pensieri corrisponde alla differenza nella struttura differenziale delle rispettive funzioni di contenuto. Il pensiero "Espero è Espero" ha una funzione di contenuto $\mathscr{C}$ la cui struttura differenziale è triviale nella direzione e: la co-estensione dei due "Espero" è insensibile a qualsiasi variazione ambientale, perché è garantita dalla struttura logica del pensiero indipendentemente dall'ambiente. Il pensiero "Espero è Fosforo" ha una funzione di contenuto la cui struttura differenziale nella direzione e è non triviale: vi sono ambienti controfattuali — mondi in cui il pianeta della sera e quello del mattino non coincidono — in cui il pensiero sarebbe falso. La derivata parziale $\partial\mathscr{C}/\partial e$ è nulla per il primo pensiero e non nulla per il secondo: questa differenza nella struttura differenziale è la controparte relazionalista della differenza cognitiva fregeana.

La superiority del trattamento relazionalista rispetto alla semantica bidimensionale standard su questo punto è sottile ma reale. La semantica bidimensionale tratta le intenzioni primarie come funzioni da mondi a estensioni, dove i mondi sono specificati in modo indipendente dai soggetti. Il relazionalismo indica che la differenza cognitiva tra i due pensieri non è soltanto modale — non riguarda soltanto quali mondi rendono il pensiero vero — ma geometrica: riguarda la *forma* della sensibilità del pensiero alle variazioni ambientali. Questa forma geometrica — la distribuzione

delle curvature di $\mathscr{C}$ — è più fine degli insiemi di mondi possibili: è una struttura differenziale che l'analisi per insiemi di mondi non può catturare perché si svolge a un livello di granularità inferiore a quello delle partizioni di W.

In altri termini: due pensieri possono avere la stessa partizione di W (la stessa intenzione secondaria) e la stessa funzione da W a W (la stessa intenzione primaria) ma differire nella struttura differenziale locale di $\mathscr{C}$ — nell'orientamento e nell'ampiezza delle curvature nella regione dell'ambiente attuale. Questa differenza differenziale è cognitivamente rilevante ma modalmente invisibile: non si manifesta in differenze di verità in mondi possibili, ma in differenze nella *struttura dell'inferenza* che il soggetto è disposto a compiere, nelle *sensibilità controfattuali locali* del pensiero, nel *modo* in cui il pensiero varia con piccole perturbazioni dell'ambiente attuale.

Il relazionalismo semantico propone dunque uno spazio del contenuto che è più fine dello spazio modale standard: uno spazio in cui la struttura differenziale locale — la geometria infinitesimale del contenuto — costituisce un livello di individuazione dei pensieri che né la semantica estensionale né quella intensionale standard, né la semantica bidimensionale nella sua versione attuale, riescono a catturare. Questo livello differenziale è il contributo tecnico più specifico che il relazionalismo offre al problema del contenuto, e costituisce il punto di maggiore originalità rispetto ai framework disponibili.

La Comunicazione come Accordo Parziale tra Geometrie Semantiche Distinte

---

XXXV. Il problema della comunicazione sotto il relazionalismo

Tra tutte le implicazioni del relazionalismo semantico, quella che interroga più direttamente la nostra esperienza quotidiana riguarda la comunicazione linguistica. Quando due parlanti usano lo stesso termine, condividono davvero il medesimo contenuto? La risposta della semantica tradizionale — sia nella versione descrittivista sia nella versione causale-esternalista — è tendenzialmente affermativa, pur con diversi gradi di qualificazione. Il relazionalismo semantico rende questa risposta problematica in modo strutturale.

Se la funzione di contenuto $\mathscr{C}$ è indicizzata al sistema cognitivo — se $\mathscr{C}$^A ≠ $\mathscr{C}$^B per sistemi cognitivi A e B con storie causali e strutture interne diverse —, allora due parlanti che usino il medesimo termine "acqua" hanno funzioni di contenuto che coincidono nei valori per gli ambienti centrali e canonici (gli ambienti in cui entrambi hanno avuto contatto diretto con esemplari d'acqua) ma possono divergere nella struttura differenziale locale e nella curvatura nella regione dei casi controfattuali e periferici. Il chimico e il bambino condividono il riferimento di "acqua" agli esemplari

centrali, ma la geometria dei loro spazi semantici nella regione intorno a quel riferimento è profondamente diversa: per il chimico, la frontiera semantica tra acqua e non-acqua è ancorata alla struttura molecolare e corre lungo confini che il bambino non ha neppure la capacità di percepire come frontiere.

Questa divergenza è, in un senso preciso, la condizione normale della comunicazione, non un'eccezione patologica. Non vi è ragione per cui due sistemi cognitivi con storie causali diverse, strutture interne diverse, competenze enciclopediche diverse, debbano avere la medesima funzione di contenuto $\mathscr{C}$ per i termini che condividono. La comunicazione riuscita non richiede identità di funzioni di contenuto: richiede qualcosa di meno — un *accordo parziale* sufficiente per i fini pratici e cognitivi della comunicazione.

Il problema teorico che si apre è: che cosa rende un accordo *sufficiente*? Qual è la struttura minima della convergenza tra $\mathscr{C}$^A e $\mathscr{C}$^B che costituisce comunicazione riuscita? E come si caratterizza formalmente il grado di accordo tra due geometrie semantiche distinte?

XXXVI. L'accordo parziale come sovrapposizione di fibre semantiche

Propongo di formalizzare l'accordo parziale tra due funzioni di contenuto come la struttura della loro *intersezione semantica* — la sottostruttura di $S_A \times E$ e di $S_B \times E$ su cui le due funzioni convergono a valori sufficientemente simili.

Definizione. Per due sistemi cognitivi A e B con funzioni di contenuto $\mathscr{C}$^A e $\mathscr{C}$^B, e per una soglia di similarità $\varepsilon > 0$, la *regione di accordo* è:

$$\Omega_\varepsilon(A, B) = \{ e \in E : \exists\ s_A \in S_A\ \exists\ s_B \in S_B : d_P(\mathscr{C}^A(s_A, e), \mathscr{C}^B(s_B, e)) < \varepsilon \}$$

dove d_P è la metrica sullo spazio delle proposizioni. La regione di accordo è l'insieme degli ambienti e in cui esistono stati interni dei due sistemi che producono contenuti sufficientemente simili. È, per così dire, il terreno comune su cui i due sistemi si incontrano semanticamente.

La comunicazione riuscita richiede che l'ambiente attuale del contesto comunicativo cada nella regione di accordo: che vi siano stati dei due sistemi che, nell'ambiente comune in cui la comunicazione si svolge, producono contenuti sufficientemente vicini. Questa condizione è più debole dell'identità delle funzioni di contenuto: non richiede che $\mathscr{C}^A = \mathscr{C}^B$ globalmente, ma che convergano localmente nella regione ambientale rilevante.

La struttura della regione di accordo rivela immediatamente perché certi tipi di comunicazione siano più solidi di altri, e certi più fragili. La comunicazione sui *termini osservativi ordinari* — colori, forme, oggetti di medio formato — ha una regione di accordo molto ampia: le funzioni di contenuto di parlanti diversi convergono su una vasta gamma di ambienti, perché le strutture percettive e cognitive di base sono largamente condivise all'interno della specie. La comunicazione sui *termini teorici specialistici* — i termini della fisica quantistica, della biologia molecolare, della giurisprudenza — ha una regione di accordo ristretta e delicata: le funzioni di contenuto di un fisico e di un non-fisico divergono significativamente già per ambienti moderatamente distanti dal canonico, e la comunicazione riuscita richiede procedure attive di calibrazione e allineamento.

Un aspetto particolarmente interessante della formalizzazione è che la regione di accordo non è necessariamente simmetrica rispetto alle direzioni dello spazio E. Due parlanti possono avere funzioni di contenuto che convergono per certi tipi di variazione ambientale e divergono per altri. La comunicazione può essere solida rispetto alle variazioni di tipo percettivo — entrambi i parlanti concordano sul colore dell'acqua, la sua viscosità, il suo odore — ma fragile rispetto alle variazioni di tipo chimico-molecolare, in cui le funzioni di contenuto hanno curvature diverse. La forma anisotropa della regione di accordo riflette la forma anisotropa delle competenze cognitive dei parlanti.

XXXVII. La comunicazione come processo di allineamento geometrico

La comunicazione non è soltanto il riconoscimento di un accordo preesistente: è spesso un *processo* che trasforma le funzioni di contenuto dei partecipanti, avvicinandole in certi aspetti e talvolta divergendole in altri. Questo processo di allineamento dinamico è un fenomeno ben documentato in psicologia cognitiva e linguistica — si parla di *conceptual alignment*, di *common ground* costruito nel corso dell'interazione, di *grounding* dei riferimenti attraverso cicli di verifica e correzione.

Nel quadro relazionalista, l'allineamento comunicativo è una *deformazione delle funzioni di contenuto* indotta dalla comunicazione stessa. Quando A trasmette un'informazione a B, e B la comprende e la integra, la funzione di contenuto di B si deforma: certe regioni dello spazio semantico si avvicinano alla struttura di $\mathscr{C}$^A, la curvatura di $\mathscr{C}$^B nella regione tematica si adatta alla curvatura di $\mathscr{C}$^A. La comunicazione riuscita è quella che produce una deformazione convergente: dopo l'interazione, la regione di accordo $\Omega_\varepsilon(A, B)$ si espande.

Formalmente, se $\mathscr{C}$^B_0 è la funzione di contenuto di B prima dell'interazione con A, e $\mathscr{C}$^B_t è quella dopo un'interazione di durata t, la comunicazione è *convergente* se:

$$\mu(\Omega_\varepsilon(A, \mathscr{C}^B_t)) > \mu(\Omega_\varepsilon(A, \mathscr{C}^B_0))$$

dove μ è una misura appropriata sulla regione di accordo. Il grado di convergenza misura quanto la comunicazione ha allineato le geometrie semantiche dei due sistemi.

Questa formalizzazione permette di distinguere tipi diversi di comunicazione in funzione della struttura della deformazione che producono. La *comunicazione ostensiva* — mostrare un oggetto indicandolo — produce una deformazione localizzata: la funzione di contenuto di B viene ancorata a un punto specifico dello spazio E (l'oggetto indicato) senza modificare la struttura globale. La *comunicazione descrittiva* — spiegare le proprietà di qualcosa — produce una deformazione più estesa: aggiunge curvatura alla funzione di contenuto di B in una regione ampia dello spazio semantico, arricchendo la struttura differenziale locale. La *comunicazione definitoria* — fornire definizioni teoriche o scientifiche — produce la deformazione più profonda: modifica l'orientamento delle curvature, riallineando le frontiere semantiche di B con quelle dell'esperto.

La divisione del lavoro linguistico burghiana trova in questa struttura una caratterizzazione più precisa. La delega semantica — il fatto che i parlanti ordinari usino termini il cui contenuto è parzialmente determinato dagli esperti — corrisponde alla struttura seguente: la funzione di contenuto del parlante ordinario ha una regione di accordo con quella dell'esperto sufficiente per gli usi comunicativi ordinari, ma la sua curvatura nella regione dei casi periferici e controfattuali segue la curvatura della funzione dell'esperto piuttosto che una struttura autonoma. Il parlante ordinario ha, per così dire, *delegato* la forma della propria curvatura semantica periferica all'esperto: la propria funzione di contenuto si estende dalla regione centrale di accordo verso la periferia seguendo il gradiente della funzione dell'esperto, anche nelle regioni in cui il parlante non ha competenza diretta.

XXXVIII. Il disaccordo comunicativo come divergenza geometrica

Il lato oscuro dell'allineamento comunicativo è il *disaccordo* — non soltanto il disaccordo esplicito su fatti, ma il disaccordo semantico latente che si produce quando due parlanti usano gli stessi termini con funzioni di contenuto strutturalmente divergenti, senza rendersene conto. Questo tipo di disaccordo — che potremmo chiamare *disaccordo geometrico occulto* — è più insidioso del disaccordo esplicito perché non si manifesta come contraddizione ma come incomunicabilità sottile, come dialogo che scorre in superficie ma non tocca mai lo stesso nucleo semantico.

I dibattiti filosofici, scientifici e politici sono costellati di esempi in cui la divergenza non è su proposizioni enunciate esplicitamente ma sulla struttura delle funzioni di contenuto dei termini chiave. Due interlocutori che discutano di "libertà", "giustizia", "razionalità" o "natura" possono usare gli stessi termini producendo proferimenti che

si accordano perfettamente su un piano superficiale — condividono la regione di accordo centrale — ma divergono radicalmente nella struttura differenziale locale: la curvatura delle funzioni di contenuto nelle regioni periferiche e controfattuali è orientata in modi opposti.

Nel quadro relazionalista, il disaccordo geometrico occulto si manifesta come segue. Siano $\mathscr{C}$^A_lib e $\mathscr{C}$^B_lib le funzioni di contenuto del termine "libertà" per i due interlocutori A e B. Nella regione degli ambienti canonici — i casi paradigmatici di libertà e di non-libertà che entrambi riconoscono — le due funzioni concordano: $\Omega_\varepsilon(A, B)$ è ampia rispetto ai casi centrali. Ma la struttura differenziale delle due funzioni nella regione dei *casi limite* — i casi controfattuali in cui la libertà di qualcuno è in conflitto con quella di qualcun altro, o in cui la libertà individuale contrasta con strutture collettive — è divergente. Le curvature di $\mathscr{C}$^A_lib e $\mathscr{C}$^B_lib hanno orientamenti opposti in questa regione: ciò che per A è una variazione interna alla categoria "libertà" (una libertà diminuita ma preservata), per B è una transizione al di là della frontiera semantica (non-libertà).

Formalmente, il disaccordo geometrico occulto in una regione $R \subseteq E$ è la divergenza dei differenziali:

$$\Delta d\mathscr{C}_R = \int_R \| d\mathscr{C}^A - d\mathscr{C}^B \|^2 \, de$$

dove la norma è calcolata rispetto alla metrica d_P sullo spazio delle proposizioni. Un grande valore di $\Delta d\mathscr{C}_R$ indica che, nella regione R degli ambienti, le strutture differenziali delle due funzioni di contenuto sono marcatamente diverse: piccole variazioni ambientali producono variazioni di contenuto in direzioni diverse per i due sistemi. Questo è il disaccordo semantico profondo — quello che nessuna quantità di accordo sui casi centrali può rivelare o risolvere.

La diagnosi del disaccordo geometrico occulto ha implicazioni metodologiche rilevanti sia per la filosofia che per le scienze sociali. Una teoria del disaccordo razionale che si limiti alla struttura delle proposizioni enunciate — che analizza il disaccordo come contraddizione tra credenze esplicite — manca la dimensione geometrica: la parte più profonda del disaccordo tra interlocutori non è nelle proposizioni che affermano ma nella forma delle funzioni di contenuto con cui le interpretano. Metodologicamente, rivelare il disaccordo geometrico richiede di esplorare sistematicamente i casi limite e controfattuali — di mappare la curvatura delle funzioni di contenuto nella regione periferica — anziché confrontare le posizioni sui casi canonici.

Questo suggerisce una tecnica di chiarificazione filosofica che il relazionalismo semantico rende formalmente motivata: la proliferazione deliberata dei casi controfattuali — la costruzione sistematica di scenari ai margini delle categorie in discussione — non è un esercizio di fantasia o di digressione, ma la procedura metodologica appropriata per esplorare la struttura differenziale delle funzioni di contenuto e identificare le regioni di divergenza geometrica. L'esperimento mentale, nella tradizione filosofica analitica, è esattamente questo: una sonda nelle regioni

periferiche dello spazio semantico, un modo di misurare la curvatura di $\mathscr{C}$ dove la regione di accordo centrale non arriva.

XXXIX. Il ruolo del disaccordo geometrico nella revisione concettuale

Il disaccordo geometrico occulto non è sempre un ostacolo alla comunicazione: è spesso il motore del progresso concettuale. Quando due sistemi cognitivi con curvature semantiche divergenti sono costretti dalla pratica comunicativa a confrontarsi sui casi limite — quando i casi controfattuali emergono nel dibattito e rivelano la divergenza delle strutture differenziali — si apre la possibilità di una *revisione concettuale*: una deformazione delle funzioni di contenuto di uno o entrambi i sistemi tale da ridurre la divergenza geometrica in regioni precedentemente problematiche.

La revisione concettuale ha una struttura diversa dall'aggiornamento bayesiano delle credenze. L'aggiornamento bayesiano modifica le probabilità di proposizioni date nuove evidenze: è un processo che opera sui valori della funzione di credenza, non sulla sua struttura. La revisione concettuale modifica la funzione di contenuto stessa — la sua curvatura, la posizione delle frontiere semantiche, l'orientamento delle sensibilità controfattuali. È una deformazione dello spazio semantico, non un aggiornamento di posizioni in uno spazio fisso.

Questa distinzione ha conseguenze per la razionalità della revisione concettuale. Se la revisione concettuale modifica la struttura dello spazio semantico, non è valutabile con i criteri della coerenza proposizionale standard: non si può dire che la revisione è razionale perché segue da premesse che il soggetto già accettava, perché le premesse appartengono al vecchio spazio semantico e le conclusioni al nuovo, e la transizione tra i due non è deduttiva ma geometrica — una deformazione continua o discontinua dello spazio. La razionalità della revisione concettuale è una questione di *coerenza geometrica*: la nuova struttura della funzione di contenuto deve essere localmente coerente (i differenziali devono integrare a funzioni ben definite), globalmente connessa (la traiettoria della revisione deve essere continua, salvo per le rotture rivoluzionarie), e adattivamente appropriata (la nuova curvatura deve allinearsi meglio con la struttura causale dell'ambiente).

Questi tre criteri — coerenza locale, connessione globale, adeguatezza adattiva — costituiscono la *norma geometrica della razionalità concettuale* nel quadro relazionalista. Non sono la stessa cosa della razionalità proposizionale standard, ma non sono neppure arbitrari: sono vincoli strutturali derivati dalla geometria dello spazio semantico e dalla sua relazione con l'ambiente. Una revisione concettuale che soddisfi questi criteri è razionale nel senso più profondo: non perché sia deducibile da premesse precedentemente accettate, ma perché produce una geometria semantica più coerente e più adeguata alla struttura causale del mondo che il soggetto abita.

Questa concezione della razionalità concettuale ha un'affinità con la proposta di Quine in "Two Dogmas of Empiricism" — l'idea che la revisione delle credenze possa toccare qualsiasi parte della rete, incluse le verità analitiche e logiche, purché produca una rete complessivamente migliore. Ma il relazionalismo semantico è più preciso di Quine riguardo alla struttura della "bontà" della rete: non si tratta di massimizzare la semplicità e il potere predittivo in astratto, ma di ottimizzare la coerenza geometrica locale e l'allineamento con le frontiere causali dell'ambiente. La metafora quiniana della rete di credenze — bidimensionale e strutturalmente omogenea — lascia il posto alla metafora dello spazio curvo: una struttura tridimensionale (almeno) la cui geometria è il riflesso, deformato e mediato dalla storia cognitiva del sistema, della struttura causale del mondo.

Le Due Obiezioni Principali: Zombie Relazionale e Contestualismo

---

XL. Il problema dello zombie relazionale riesaminato

Nella Parte Quarta di questa monografia il problema dello zombie relazionale è stato introdotto e lasciato aperto con dichiarata onestà. È tempo di affrontarlo con la sistematicità che merita, perché è l'obiezione più profonda che il relazionalismo semantico deve fronteggiare — quella che mette in questione non un dettaglio del formalismo ma la pretesa dell'intera impostazione di dire qualcosa di genuino sul contenuto mentale, piuttosto che sulla sua simulazione funzionale.

Lo zombie relazionale è un sistema che, per ogni coppia $\langle s, e \rangle \in S \times E$, produce la stessa funzione di contenuto $\mathscr{C}$ di un sistema cognitivo con comprensione genuina, ma in assenza di qualsiasi intenzionalità originaria o esperienza semantica. L'obiezione si articola in due passi. Il primo: se il relazionalismo definisce il contenuto come struttura differenziale di $\mathscr{C}$ — come proprietà geometrica di una funzione matematica — allora uno zombie relazionale è concettualmente possibile, perché qualsiasi funzione matematica può in linea di principio essere computata da sistemi privi di comprensione. Il secondo: se lo zombie relazionale è possibile, il relazionalismo non offre condizioni sufficienti per il contenuto mentale genuino, ma soltanto per la sua controparte funzionale. Torna, in forma nuova, la critica searliana alla Stanza Cinese: la struttura formale non è la semantica.

La risposta che intendo sviluppare non è una refutazione diretta della possibilità dello zombie relazionale — tentativo che sarebbe presumibilmente fallace — ma è una *ridefinizione del peso teorico* di questa possibilità. Argomenterò che l'obiezione, per essere cogente contro il relazionalismo, deve assumere una nozione di "contenuto mentale genuino" che è indipendente da qualsiasi struttura relazionale e causale — una nozione che è, in ultima istanza, o indimostrabile o circolare. In assenza di questa

nozione indipendente, il relazionalismo non è confutato dallo zombie relazionale: è semplicemente sfidato a specificare quale tipo di struttura causale costituisce contenuto genuino, il che è esattamente il programma di ricerca che la proposta intende promuovere.

Il nucleo dell'argomentazione è il seguente. Quando si dice che lo zombie relazionale "computa $\mathscr{C}$ senza capire nulla", si presuppone che vi sia un fatto indipendente dalla struttura di $\mathscr{C}$ che determina se il sistema capisce o non capisce. Ma quale fatto potrebbe essere? Non può essere un fatto intrinseco agli stati interni del sistema — questa mossa sarebbe internalista, e l'internalismo ha le difficoltà già discusse. Non può essere un fatto puramente relazionale diverso da $\mathscr{C}$ — perché allora il relazionalismo è semplicemente incompleto, non sbagliato, e il compito è identificare la struttura relazionale aggiuntiva rilevante. Non può essere un fatto fenomenico bruto — un quale irriducibile della comprensione — perché questa mossa presuppone risolta la questione filosofica più difficile di tutte, quella del rapporto tra struttura funzionale e esperienza cosciente.

Il relazionalismo semantico risponde dunque all'obiezione dello zombie non negandone la forza ma chiarendo che essa sposta il problema da *dove si trova il contenuto* a *quale tipo di processo causale costituisce contenuto autentico piuttosto che mera simulazione*. Questa è una domanda genuina — la domanda più profonda della filosofia della mente — ma non è una domanda che il relazionalismo semantico, come teoria del contenuto, sia tenuto a risolvere nella sua interezza. Il relazionalismo fornisce le condizioni strutturali del contenuto — la forma che una funzione $\mathscr{C}$ deve avere perché uno stato mentale abbia contenuto — ma non pretende di fornire le condizioni fisiche ultime della realizzazione di questa struttura in sistemi che comprendono genuinamente.

È il medesimo tipo di risposta che la termodinamica dà all'obiezione: "ma il calore è davvero energia cinetica media, oppure vi è qualcosa in più nell'esperienza del caldo che la termodinamica non cattura?" La termodinamica risponde: ciò che compete a una teoria del calore come fenomeno fisico è specificarne la struttura causale e le leggi che la governano; la questione di come quella struttura fisica si colleghi all'esperienza fenomenica è una questione separata, che la termodinamica non pretende di risolvere ma che neppure la confuta. Analogamente, il relazionalismo semantico si occupa della struttura del contenuto come fenomeno cognitivo-semantico; la questione di come questa struttura si colleghi alla comprensione fenomenica è una questione separata che appartiene al problema difficile della coscienza, non alla teoria del contenuto.

Questa risposta ha però un costo che è onesto riconoscere: lascia aperta la possibilità che il relazionalismo semantico — per quanto adeguato come teoria della *struttura* del contenuto — sia incompleto come teoria del *contenuto tout court*, se si ritiene che il contenuto richieda necessariamente una dimensione fenomenica irriducibile. Chi accetta questa premessa non troverà nel relazionalismo una risposta soddisfacente. Chi invece ritiene — come chi scrive è inclinato a ritenere, pur con incertezza — che il problema fenomenico del contenuto sia ortogonale al problema strutturale, troverà nel

relazionalismo una proposta che risolve la parte strutturale del problema con sufficiente precisione da costituire un avanzamento genuino.

Un'ultima precisazione. La distinzione tra equivalenza estensionale e intensionale della funzione di contenuto, introdotta nella Parte Quarta, offre una risorsa aggiuntiva per rispondere allo zombie relazionale. Lo zombie computa la medesima funzione in senso estensionale — produce gli stessi output per gli stessi input — ma non in senso intensionale: il tipo di processo causale attraverso cui la funzione è prodotta è diverso. Se si sostiene — come il relazionalismo implica — che il contenuto genuino richiede non solo la funzione ma il tipo causale del processo che la genera, allora lo zombie relazionale non ha il medesimo contenuto ma soltanto la medesima mappa input-output. Questo non risolve il problema — sposta soltanto la questione su quale tipo causale di processo sia necessario — ma chiarisce che la sfida dello zombie non è una refutazione del relazionalismo ma un programma di ricerca sulla natura dei processi causali costitutivi del contenuto.

XLI. L'obiezione del contestualismo: il relazionalismo è una forma sofisticata di relativismo semantico?

La seconda obiezione principale al relazionalismo semantico è strutturalmente diversa dalla prima: non proviene dal problema difficile della coscienza ma dalla semantica formale e dalla filosofia del linguaggio. L'obiezione è che il relazionalismo, nella misura in cui rende il contenuto dipendente dalla coppia $\langle s, e \rangle$ — dallo stato interno del soggetto congiunto all'ambiente —, collassi in una forma di *contestualismo semantico* e diventi così vulnerabile alle obiezioni standard contro quella posizione.

Il contestualismo semantico — nelle sue versioni più elaborate, come quella di Recanati o di Carston — sostiene che il contenuto delle espressioni linguistiche non è determinato dalla sola semantica composizionale ma richiede arricchimento pragmatico contestuale: ciò che una frase esprime dipende dal contesto di proferimento in modo che eccede le istruzioni semantiche convenzionali. Questa posizione è stata criticata per due ragioni principali. La prima è la *perdita della stabilità comunicativa*: se il contenuto varia con il contesto, non vi è un oggetto semantico stabile che i parlanti condividono, e la comunicazione diventa inspiegabilmente efficace. La seconda è il *problema della proliferazione contestuale*: se il contesto può modificare il contenuto in modo non sistematico, qualsiasi aggiustamento ad hoc diventa teoricamente lecito, e la semantica perde il suo carattere nomologico.

Il relazionalismo semantico deve rispondere a queste due critiche, poiché la dipendenza di $\mathscr{C}$ dall'ambiente e lo espone apparentemente ad entrambe.

Alla prima critica — la perdita di stabilità comunicativa — il relazionalismo risponde attraverso la nozione di regione di accordo $\Omega_\varepsilon(A, B)$ sviluppata nella Parte Decima.

La comunicazione non richiede che due soggetti condividano la medesima funzione di contenuto: richiede che la regione di accordo sia sufficientemente ampia rispetto agli ambienti rilevanti per la comunicazione in corso. La stabilità della comunicazione è una proprietà della *topologia della regione di accordo* — della sua ampiezza, connessione e robustezza alle piccole variazioni — non dell'identità delle funzioni di contenuto. Questa stabilità è genuina: non è postulata per definizione ma emerge dalla struttura geometrica dello spazio semantico condiviso dalla comunità linguistica. La divisione del lavoro semantico, le pratiche normative, l'addestramento comune su esemplari condivisi — tutti questi meccanismi producono convergenza nelle funzioni di contenuto nelle regioni centrali, garantendo la stabilità comunicativa ordinaria senza richiedere identità globale.

Alla seconda critica — la proliferazione contestuale — il relazionalismo risponde attraverso la struttura *vincolata* della funzione $\mathscr{C}$. Non è vero che qualsiasi dipendenza dall'ambiente sia ammissibile: la funzione $\mathscr{C}$ è soggetta a vincoli strutturali precisi. Deve essere differenziabile — le variazioni di contenuto devono essere continue rispetto alle variazioni di $\langle s, e \rangle$, salvo discontinuità strutturalmente motivate. Deve essere coerente con la storia causale del sistema — le curvature di $\mathscr{C}$ sono determinate dalla struttura causale delle interazioni tra il sistema e l'ambiente, non da aggiustamenti arbitrari. Deve essere compositiva nel senso della Parte Settima — la struttura funtoriale della composizione pone vincoli sulla forma delle funzioni componenti che non possono essere violati senza produrre incoerenza geometrica. Questi vincoli escludono la proliferazione arbitraria del contenuto: la dipendenza contestuale di $\mathscr{C}$ è sistematica e strutturalmente determinata, non ad hoc.

La differenza decisiva tra il relazionalismo semantico e il contestualismo standard è questa: il contestualismo ammette dipendenza contestuale al livello dei *valori* del contenuto — ciò che un'espressione dice in un contesto varia con il contesto — senza specificare la struttura di questa dipendenza. Il relazionalismo opera al livello della *struttura della funzione di contenuto* — specifica come la dipendenza contestuale è organizzata, quali dimensioni dello spazio ambientale sono semanticamente rilevanti, quale forma ha la curvatura di $\mathscr{C}$ nelle diverse regioni. Non è la variazione del contenuto con il contesto che distingue il relazionalismo dal contestualismo, ma la struttura geometrica di quella variazione.

In termini più formali: il contestualismo afferma che il contenuto è una funzione dei contesti, senza specificare la struttura di questa funzione. Il relazionalismo afferma che questa funzione è $\mathscr{C} : S \times E \to P$, è differenziabile, ha una struttura di curvature determinata causalmente, è composizionale in senso funtoriale. La struttura è ciò che impedisce la proliferazione arbitraria: non tutte le funzioni da $S \times E$ a P sono funzioni di contenuto ammissibili, ma soltanto quelle che soddisfano i vincoli geometrici derivati dalla natura dei sistemi cognitivi e delle loro interazioni con l'ambiente.

Vi è ancora un'obiezione residua che merita risposta: se la struttura di $\mathscr{C}$ è determinata causalmente dalla storia del sistema, e sistemi con storie diverse hanno funzioni $\mathscr{C}$ diverse, come si esclude che parlanti con storie sufficientemente diverse non abbiano, in ultima istanza, significati incommensurabili — talmente divergenti nelle curvature

da non condividere neppure una regione di accordo minimale? Questa è la versione relativista più acuta dell'obiezione.

La risposta è empirica piuttosto che a priori: i sistemi cognitivi della specie umana condividono, per effetto di una storia evolutiva comune, una struttura cognitiva di base — sistemi percettivi, capacità inferenziali, disposizioni categoriali — che produce convergenza nelle funzioni di contenuto nelle regioni centrali dello spazio semantico. L'incommensurabilità radicale è concettualmente possibile ma empiricamente remota, almeno all'interno della specie: richiederebbe storie causali così divergenti da non condividere neppure gli esemplari percettivi fondamentali. Tra sistemi cognitivi di specie molto diverse — o tra sistemi artificiali con architetture radicalmente eterogenee rispetto a quella biologica — l'incommensurabilità parziale è plausibile e il relazionalismo ha gli strumenti formali per misurarla (la regione di accordo Ω_ε), non soltanto per nominarla.

Il relazionalismo semantico non nega dunque la forza del relativismo semantico come posizione limite: riconosce che l'accordo semantico tra sistemi cognitivi è una questione di grado, misurabile geometricamente, non una questione binaria di identità o differenza. Ma trasforma questa ammissione da una concessione scettica in un programma di ricerca positivo: la misurazione della regione di accordo, la caratterizzazione delle sue asimmetrie e delle sue variazioni in funzione del tipo di contenuto e del tipo di sistema cognitivo, è un'impresa teorica ed empirica genuina — e il relazionalismo semantico ne fornisce le coordinate concettuali.

Appendice Formale — Definizioni, Strutture e Vincoli del Relazionalismo Semantico

---

A.1 Strutture di base

Siano dati i seguenti spazi primitivi, assunti come insiemi topologici con struttura differenziale appropriata:

S — lo *spazio degli stati interni* di un sistema cognitivo A. Un elemento $s \in S$ rappresenta uno stato interno del sistema al livello di descrizione pertinente (neurale, funzionale, computazionale). S è assunto come varietà differenziale di dimensione finita o infinita, secondo il livello di astrazione scelto.

E — lo *spazio degli ambienti*. Un elemento $e \in E$ rappresenta una situazione ambientale possibile, intesa come struttura che include proprietà fisiche, chimiche, biologiche,

sociali e normative rilevanti per il sistema cognitivo. E ha una decomposizione naturale:

$$E = E_T \times E_P \times E_N$$

dove E_T è la componente *tipologica* (le proprietà strutturali dell'ambiente: composizione molecolare, struttura causale, tipi naturali presenti), E_P è la componente *posizionale* (la posizione e l'identità del sistema nell'ambiente: coordinate spaziotemporali, identità come nodo nella rete causale), e E_N è la componente *normativa* (le norme linguistiche, le pratiche semantiche e le autorità epistemiche della comunità in cui il sistema è inserito).

P — lo *spazio delle proposizioni*. P è equipaggiato con una metrica d_P che cattura la similarità proposizionale. La struttura esatta di P dipende dalla teoria proposizionale assunta (insiemi di mondi possibili, strutture russelleane, proposizioni neo-fregeane) e il relazionalismo è, nella misura del possibile, neutro rispetto a questa scelta.

A.2 La funzione di contenuto

Definizione A.1 (Funzione di contenuto). Una *funzione di contenuto* per il sistema A è una mappa differenziabile:

$$\mathscr{C}_A : S \times E \to P$$

tale che $\mathscr{C}_A(s, e)$ è il contenuto dello stato mentale s del sistema A quando A è inserito nell'ambiente e. La differenziabilità richiede che la mappa sia continua e che le derivate parziali di primo ordine esistano e siano continue su $S \times E$ (salvo un insieme di misura nulla corrispondente a discontinuità strutturalmente motivate — rotture semantiche).

Definizione A.2 (Differenziale del contenuto). Il *differenziale totale* di $\mathscr{C}_A$ in un punto (s_0, e_0) è:

$$d\mathscr{C}_A|_{(s_0,e_0)} = (\partial\mathscr{C}_A/\partial s)|_{(s_0,e_0)} \cdot ds + (\partial\mathscr{C}_A/\partial e)|_{(s_0,e_0)} \cdot de$$

Il termine $\partial\mathscr{C}_A/\partial s$ cattura la sensibilità del contenuto alle variazioni degli stati interni (la componente "internalista" della struttura semantica); il termine $\partial\mathscr{C}_A/\partial e$ cattura la sensibilità alle variazioni ambientali (la componente "esternalista").

Definizione A.3 (Curvatura semantica). La *curvatura semantica* di $\mathscr{C}_A$ in (s_0, e_0) lungo la direzione $v \in T(S \times E)$ è:

$$\varkappa_A(s_0, e_0, v) = d^2\mathscr{C}_A|_{(s_0,e_0)}[v, v]$$

dove $d^2\mathscr{C}^{\wedge}A$ è la forma bilineare della derivata seconda (la hessiana di $\mathscr{C}^{\wedge}A$). Alta curvatura nella direzione ambientale indica alta sensibilità del contenuto a quella dimensione dell'ambiente; alta curvatura nella direzione interna indica alta sensibilità alle variazioni degli stati interni.

A.3 Casi limite: internalismo ed esternalismo come condizioni su $\mathscr{C}^{\wedge}A$

Proposizione A.1 (Internalismo come annullamento della derivata ambientale). Il sistema A soddisfa la tesi internalista (il contenuto sopravviene sugli stati interni) se e solo se:

$\partial\mathscr{C}^{\wedge}A/\partial e = 0$ uniformemente su $S \times E$

In questo caso $\mathscr{C}^{\wedge}A$ si riduce a una funzione $\mathscr{C}^{\wedge}A : S \to P$ indipendente da E.

Proposizione A.2 (Esternalismo putnamiano come sensibilità alla componente tipologica). L'esternalismo dei tipi naturali di Putnam corrisponde alla condizione:

$\partial\mathscr{C}^{\wedge}A/\partial e_T \neq 0$ nella regione semanticamente rilevante di $S \times E_T$

con l'ulteriore condizione che la variazione sia prodotta dalla struttura causale di E_T (la chimica, la fisica) e non da convenzioni sociali (E_N).

Proposizione A.3 (Anti-individualismo burghiano come sensibilità alla componente normativa). L'anti-individualismo di Burge corrisponde alla condizione:

$\partial\mathscr{C}^{\wedge}A/\partial e_N \neq 0$ nella regione rilevante

con la variazione prodotta dalle norme semantiche della comunità (E_N) indipendentemente dalla struttura fisico-chimica dell'ambiente (E_T).

A.4 Separabilità semantica

Definizione A.4 (s-separabilità). $\mathscr{C}^{\wedge}A$ è *s-separabile* rispetto a una partizione $E = E_1 \cup E_2$ se:

$\forall s \in S \; \forall e_1 \in E_1 \; \forall e_2 \in E_2 : \mathscr{C}^{\wedge}A(s, e_1) = \mathscr{C}^{\wedge}A(s, e_2)$

La s-separabilità rispetto a una partizione dell'ambiente esprime il fatto che il sistema non discrimina semanticamente tra le due regioni ambientali: il contenuto è insensibile a quella distinzione.

Definizione A.5 (e-separabilità). $\mathscr{C}\hat{}A$ è *e-separabile* rispetto a una partizione $S = S_1 \cup S_2$ se:

$$\forall\, e \in E\ \forall\, s_1 \in S_1\ \forall\, s_2 \in S_2 : \mathscr{C}\hat{}A(s_1, e) = \mathscr{C}\hat{}A(s_2, e)$$

La e-separabilità rispetto a una partizione degli stati interni esprime il fatto che quella distinzione tra stati interni non produce differenza di contenuto: stati nella stessa partizione hanno il medesimo contenuto in ogni ambiente.

Corollario A.1 (Terra Gemella come violazione della s-separabilità). Siano E_T (Terra) e E_TG (Terra Gemella) due ambienti tali che $H_2O \in$ E_T e XYZ $\in$ E_TG. Il risultato di Putnam afferma che $\mathscr{C}\hat{}A$ non è s-separabile rispetto alla partizione {E_T, E_TG}: per lo stato s_0 condiviso da Oscar e Oscar*, $\mathscr{C}\hat{}A(s_0, E_T) \neq \mathscr{C}\hat{}A(s_0, E_TG)$.

A.5 Equivalenza psicologica e leggi psicologiche

Definizione A.6 (Equivalenza diffeomorfica). Due punti (s_1, e_1) e $(s_2, e_2) \in S \times E$ sono *psicologicamente equivalenti* se esiste un diffeomorfismo locale $\varphi : U_1 \to U_2$ tra intorni aperti dei due punti tale che:

$$\mathscr{C}\hat{}A \circ \varphi = \mathscr{C}\hat{}A \quad \text{su } U_1$$

L'equivalenza diffeomorfica cattura l'identità di *struttura semantica locale*: i due punti producono contenuti della stessa forma geometrica, anche se di valori diversi.

Tesi A.1 (Leggi psicologiche come leggi su classi di equivalenza). Le leggi psicologiche sono generalizzazioni che quantificano universalmente sulle classi di equivalenza diffeomorfica di $\mathscr{C}\hat{}A$, non sui valori di $\mathscr{C}\hat{}A$ né sugli stati interni di S. Una legge psicologica L è della forma:

$$\forall\, [(s, e)] \in [S \times E]/\sim : \Phi([(s, e)]) \to \Psi([(s, e)])$$

dove $[(s, e)]$ denota la classe di equivalenza diffeomorfica di (s, e), e Φ, Ψ sono condizioni sulla struttura locale di $\mathscr{C}\hat{}A$.

A.6 Composizione funtoriale

Definizione A.7 (Spazio degli stati compositi). Per due stati componenti con spazi S_α e S_β, lo *spazio degli stati compositi* è il fibrato:

$$S_{[\alpha,\beta]} = \{ (s_\alpha, s_\beta, \varrho) : s_\alpha \in S_\alpha, s_\beta \in S_\beta, \varrho \in R(s_\alpha, s_\beta) \}$$

dove R(s_α, s_β) è lo spazio delle relazioni sintattico-semantiche ammissibili tra s_α e s_β (applicazione funzionale, predicazione, modificazione, ecc.).

Definizione A.8 (Funzione di contenuto composita). La funzione di contenuto del pensiero composto [α, β] è:

$$\mathscr{C}_{[\alpha,\beta]}(s_\alpha, s_\beta, \varrho, e) = \Phi_\varrho(\mathscr{C}_\alpha(s_\alpha, e), \mathscr{C}_\beta(s_\beta, e))$$

dove Φ_ϱ è l'operazione semantica determinata dalla relazione sintattica ϱ.

Tesi A.2 (Composizionalità come proprietà delle strutture differenziali). La composizionalità relazionale afferma che la struttura differenziale di $\mathscr{C}_{[\alpha,\beta]}$ è determinata dalle strutture differenziali di $\mathscr{C}_\alpha$ e $\mathscr{C}_\beta$ lungo le fibre sintattiche di R(s_α, s_β). Non richiede che i valori di $\mathscr{C}_\alpha$ e $\mathscr{C}_\beta$ siano fissi indipendentemente dal contesto composizionale.

A.7 Regione di accordo e comunicazione

Definizione A.9 (Regione di accordo). Per due sistemi A e B e una soglia ε > 0, la *regione di accordo* è:

$$\Omega_\varepsilon(A, B) = \{ e \in E : \exists\, s_A \in S_A \ \exists\, s_B \in S_B : d_P(\mathscr{C}\hat{}A(s_A, e), \mathscr{C}\hat{}B(s_B, e)) < \varepsilon \}$$

Definizione A.10 (Comunicazione convergente). Una comunicazione tra A e B di durata t è *convergente* se:

$$\mu(\Omega_\varepsilon(A, \mathscr{C}\hat{}B_t)) > \mu(\Omega_\varepsilon(A, \mathscr{C}\hat{}B_0))$$

dove μ è una misura su E e $\mathscr{C}\hat{}B_t$ è la funzione di contenuto di B dopo l'interazione.

Definizione A.11 (Disaccordo geometrico). Il *disaccordo geometrico* tra A e B nella regione R ⊆ E è:

$$\Delta d\mathscr{C}_R(A, B) = \int_R \| d\mathscr{C}\hat{}A - d\mathscr{C}\hat{}B \|^2_P \, de$$

dove la norma $\|\cdot\|_P$ è indotta dalla metrica d_P su P. Un valore alto di $\Delta d\mathscr{C}_R$ indica divergenza strutturale delle funzioni di contenuto nella regione R, anche in presenza di accordo sui valori centrali.

A.8 Identità diacronica del contenuto

Definizione A.12 (Continuità della traiettoria semantica). Una traiettoria $\{(s(t), e(t)) : t \in [0, T]\}$ in $S \times E$ è *semanticamente continua* se la funzione composita $\mathscr{C}_A(s(t), e(t))$ è continua in t. L'identità diacronica del contenuto di un pensiero è preservata lungo traiettorie semanticamente continue.

Definizione A.13 (Rottura semantica). Una *rottura semantica* in $t_0 \in (0, T)$ è una discontinuità di $\mathscr{C}_A(s(t), e(t))$ in t_0 che non è rimovibile per nessuna reparametrizzazione della traiettoria. Le rotture semantiche corrispondono a rivoluzioni concettuali nel senso kuhniano.

Appendice Bibliografica

La seguente selezione bibliografica non è esaustiva ma orienta il lettore verso i testi che hanno strutturato il problema affrontato in questa monografia e verso quelli con cui il relazionalismo semantico si trova in dialogo diretto. I testi sono raggruppati tematicamente, non per ordine alfabetico né cronologico, perché la struttura tematica riflette più fedelmente l'architettura argomentativa della ricerca.

Sul problema del contenuto e la sua collocazione: testi fondativi

Hilary Putnam, *The Meaning of "Meaning"*, in *Mind, Language and Reality: Philosophical Papers*, vol. 2, Cambridge University Press, Cambridge 1975. Il testo che introduce l'esperimento della Terra Gemella e il verdetto "meanings ain't in the head". Qualsiasi riflessione sul contenuto mentale deve confrontarsi con questo saggio prima di tutto.

Saul A. Kripke, *Naming and Necessity*, Harvard University Press, Cambridge (MA) 1980. La teoria causale del riferimento e la distinzione tra necessità metafisica e necessità epistemica, che condiziona ogni discussione successiva sui tipi naturali e sulla rigidità del riferimento.

Tyler Burge, *Individualism and the Mind*, in «Midwest Studies in Philosophy», vol. 4, 1979, pp. 73-121. L'anti-individualismo sociale, l'esperimento dell'artrite, la tesi che il contenuto mentale sia co-determinato dalle norme semantiche della comunità.

Jerry A. Fodor, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, MIT Press, Cambridge (MA) 1987. La risposta internalista più sistematica: la distinzione

stretto/largo, il contenuto stretto come condizione per le leggi psicologiche, la proposta dei mondi nozionali.

Jerry A. Fodor, *A Theory of Content and Other Essays*, MIT Press, Cambridge (MA) 1990. Sviluppo e revisione delle posizioni di *Psychosemantics*, con la teoria informazionale asimmetrica della dipendenza.

Sulla semantica dei mondi possibili e la struttura bidimensionale

David Kaplan, *Demonstratives*, in Joseph Almog, John Perry, Howard Wettstein (a cura di), *Themes from Kaplan*, Oxford University Press, Oxford 1989. Il sistema formale per gli indicali e la distinzione carattere/contenuto, indispensabile per la Parte Nona di questa monografia.

David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, Oxford 1996. La semantica bidimensionale e la sua motivazione nella teoria della coscienza; il trattamento della necessità a posteriori in termini di intenzioni primarie e secondarie.

David J. Chalmers, *Does Conceivability Entail Possibility?*, in Tamar Szabo Gendler, John Hawthorne (a cura di), *Conceivability and Possibility*, Oxford University Press, Oxford 2002. Il nesso tra possibilità primaria e concezione ideale; cruciale per il confronto sviluppato nella Parte Nona.

Martin Davies, Lloyd Humberstone, *Two Notions of Necessity*, in «Philosophical Studies», vol. 38, 1980, pp. 1-30. Il saggio originario della semantica bidimensionale, anteriore all'elaborazione chalmeriana.

Sul riferimento singolare e i pensieri de se

Gareth Evans, *The Varieties of Reference*, Oxford University Press, Oxford 1982. La teoria del pensiero singolare come pensiero che richiede contatto informazionale con l'oggetto; la tesi di trasparenza discussa nella Parte Ottava.

John Perry, *The Problem of the Essential Indexical*, in «Noûs», vol. 13, 1979, pp. 3-21. Il caso del carrello del supermercato e la irriducibilità dei pensieri de se; punto di partenza per la trattazione degli indicali nella Parte Quinta.

John Perry, *Thought without Representation*, in «Proceedings of the Aristotelian Society», Supplementary Volume, vol. 60, 1986, pp. 137-151. Lo sviluppo della teoria dei pensieri "senza oggetto" in caso di fallimento del riferimento; rilevante per il problema del riferimento singolare.

Sulla teleosemantica

Ruth Millikan, *Language, Thought, and Other Biological Categories: New Foundations for Realism*, MIT Press, Cambridge (MA) 1984. L'opera fondativa della teleosemantica millikaniana: la funzione propria come determinante del contenuto.

Fred Dretske, *Explaining Behavior: Reasons in a World of Causes*, MIT Press, Cambridge (MA) 1988. La versione informazionale-teleologica del contenuto: il contenuto come ciò per cui uno stato mentale è stato selezionato in quanto segnale.

Karen Neander, *Misrepresentation and Malfunction*, in «Philosophical Studies», vol. 79, 1995, pp. 109-141. Il problema della malrappresentazione nella teleosemantica e le risorse per affrontarlo; critico per la valutazione sviluppata nella Parte Quarta.

Sul problema dell'esclusione causale e l'efficacia mentale

Jaegwon Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, MIT Press, Cambridge (MA) 1998. L'argomento dell'esclusione causale nella sua forma più compiuta; sfida diretta a qualsiasi teoria che attribuisca al contenuto largo efficacia causale.

Stephen Yablo, *Mental Causation*, in «Philosophical Review», vol. 101, 1992, pp. 245-280. La risposta alla sfida di Kim attraverso la nozione di causazione come realizzazione parziale; rilevante per la Parte Seconda.

Sulla cognizione estesa e i confini del sistema cognitivo

Andy Clark, David Chalmers, *The Extended Mind*, in «Analysis», vol. 58, 1998, pp. 7-19. Il principio di parità e la proposta che il sistema cognitivo si estenda fisicamente al di là del cranio; discusso nella Parte Seconda come evidenza della circolarità del dualismo di dominio.

Andy Clark, *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford University Press, Oxford 2008. Sviluppo sistematico della tesi della cognizione estesa con risposta alle obiezioni; importante per la sezione sui sistemi distribuiti nella Parte Sesta.

Sulla semantica contestuale e la pragmatica del contenuto

François Recanati, *Literal Meaning*, Cambridge University Press, Cambridge 2004. Il contestualismo radicale e la modulazione del contenuto per via pragmatica; posizione rispetto alla quale il relazionalismo semantico si definisce per differenza nella Parte Undicesima.

Robyn Carston, *Thoughts and Utterances: The Pragmatics of Explicit Communication*, Blackwell, Oxford 2002. La teoria della rilevanza e l'arricchimento esplicito del contenuto; rilevante per la distinzione tra contenuto semantico e contenuto pragmatico.

Sull'autoconoscenza e il privilegio in prima persona

Tyler Burge, *Our Entitlement to Self-Knowledge*, in «Proceedings of the Aristotelian Society», vol. 96, 1996, pp. 91-116. L'esternalismo sull'autoconoscenza: come l'anti-individualismo si applica alla conoscenza che il soggetto ha dei propri stati mentali.

Brie Gertler (a cura di), *Privileged Access: Philosophical Accounts of Self-Knowledge*, Ashgate, Aldershot 2003. Raccolta di testi fondamentali sul dibattito tra internalismo ed esternalismo epistemico sull'autoconoscenza; utile come panorama del problema affrontato nella Parte Ottava.

Sul cambiamento concettuale e la razionalità della revisione

W.V.O. Quine, *Two Dogmas of Empiricism*, in «Philosophical Review», vol. 60, 1951, pp. 20-43. Il fallibilismo semantico e la possibilità di revisione di qualsiasi enunciato, incluse le verità analitiche; posizione rispetto alla quale la norma geometrica della razionalità concettuale si definisce per affinamento nella Parte Decima.

Thomas S. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago 1962. Le rivoluzioni scientifiche come rotture semantiche e il problema dell'incommensurabilità; il quadro relazionalista formalizza la nozione kuhniana di incommensurabilità come discontinuità topologica della traiettoria semantica.

Paul Feyerabend, *Against Method: Outline of an Anarchistic Theory of Knowledge*, New Left Books, London 1975. La versione più radicale dell'incommensurabilità; utile come termine di confronto per la posizione più strutturata del relazionalismo.

Per un orientamento in geometria differenziale applicata alla filosofia

Michael Spivak, *A Comprehensive Introduction to Differential Geometry*, 5 voll., Publish or Perish, Houston 1979. Riferimento tecnico standard per la geometria differenziale usata nel formalismo di questa monografia; i concetti di varietà differenziale, fibrato, curvatura e differenziale sono trattati con la precisione necessaria.

René Thom, *Structural Stability and Morphogenesis: An Outline of a General Theory of Models*, Benjamin, Reading (MA) 1975. L'applicazione della teoria delle catastrofi ai fenomeni

morfologici e semantici; ispira, pur senza adottarne i dettagli tecnici, la nozione di rottura semantica come catastrofe nella traiettoria di *C*.